

EAST Search History

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
L1	2442	flavivirus	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT	OR	ON	2006/07/09 20:33
L2	7135	envelope adj protein	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT	OR	ON	2006/07/09 20:33
L3	681	I1 and I2	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT	OR	ON	2006/07/09 20:33
L4	1473	domain adj III	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT	OR	ON	2006/07/09 20:34
L5	37	I3 and I4	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT	OR	ON	2006/07/09 20:34
L6	1065793	inhibit\$	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT	OR	ON	2006/07/09 20:34
L7	31	I5 and I6	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT	OR	ON	2006/07/09 20:34
L8	246025	antibod\$	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT	OR	ON	2006/07/09 20:34
L9	31	I7 and I8	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT	OR	ON	2006/07/09 20:35

EAST Search History

L10	168742	ligand	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT	OR	ON	2006/07/09 20:35
L11	17	I9 and I10	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT	OR	ON	2006/07/09 20:35



A service of the National Library of Medicine
and the National Institutes of Health

My NCBI
[Sign In] [Register]

[All Databases](#)[PubMed](#)[Nucleotide](#)[Protein](#)[Genome](#)[Structure](#)[OMIM](#)[PMC](#)[Journals](#)[Books](#)Search **PubMed**for **flavivirus domain III**[Preview](#)[Go](#)[Clear](#)[Limits](#)[Preview/Index](#)**[History](#)**[Clipboard](#)[Details](#)[About Entrez](#)[Text Version](#)[Entrez PubMed](#)[Overview](#)[Help | FAQ](#)[Tutorials](#)[New/Noteworthy](#) [E-Utilities](#)**Search****Most Recent Queries****Time Result**[#24](#) Search **flavivirus domain III**20:39:18 [58](#)[#23](#) Search **flavivirus**20:39:15 [8529](#)[Clear History](#)**PubMed Services**[Journals Database](#)[MeSH Database](#)[Single Citation Matcher](#)[Batch Citation Matcher](#)[Clinical Queries](#)[Special Queries](#)[LinkOut](#)[My NCBI](#)**Related Resources**[Order Documents](#)[NLM Mobile](#)[NLM Catalog](#)[NLM Gateway](#)[TOXNET](#)[Consumer Health](#)[Clinical Alerts](#)[ClinicalTrials.gov](#)[PubMed Central](#)

Jul 6 2006 07:09:19

77e

FAPELANNTFVIDGPETEECTANRAWNSMEVEDFGFGLTSTRMFLRIRETNTTECD
KIIGTAVKNNMAVHSDLSYWIESGLNDTWKLERAVLGEVKSTWPEHTLWGDGVLES
DLIIPITLAGPRSNHNRPGYKTQNGQPWDEGRVEIDFDYCPGTTVTISDSCGHRGPA
ARTTTESGKLITDWCCRSTLPPLRFQTENGCWYGMIEIRPTRHDEKTLVQSRVNAYNA
DMIDPFQGLLVVFLATQEVLRKRWTAKISIPAIMLALLVLVFGGITYTDVLRVILV
GAFAEANSGGDVVHLALMATFKIQPVFLVASFLKARWTNQESILLMLAAAFFQMAYY
DAKNVLSWEVPDVLNLSVAWMILRAISFTNTSNVVVPLLALLTPGLKCLNLDVYRIL
LLMVGVGSLIKEKRSSAAKKKGACLICLALASTGVFNPMILAAGLMACDPNRKRGWPA
TEVMTAVGLMFAIVGGLAELDIDSMAPMTIAGLMFVAFVISGKSTDMWIERTADITW
ESDAEITGSSERVDVRLDDGDNFQLMNDPGAPWKIWMMLRMACLAISAYTPWAILPSVI
GFWITLQYTKRGGVLWDTPSPKEYKKGDTTGTGYRIMTRGLLGSYQAGAGVMVEGVFH
TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGHDEVQMI VVEPGKN
VKNVQTKPGVFKTPEGEIGAVTLDYPTGTSGSPIVDKNGDVIGLYGNVIMPNGSYIS
AIVQGERMEEPAPAGFEPEMLRKKQITVLDLHPGAGKTRKILPQIIKEAINKRLRTAV
LAPTRVVAEMSEALRGLPIRYQTSAVHREHSGNEIVDMCHATLTHRLMSPHRPNY
NLFIMDEAHFTDPASIAARGYIATKVELGEAAAFMTATPPGTSDFPESNAPITDMQ
TEIPDRAWNTGYEWITEYVGKTVWFVPSVKMGNEIALCLQRAGKKVIQLNRKSYET
PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIIIEEGDGRVILGEP
SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLVAQ
LYQPEREKVYTM DGEYRLRGEERKNFLEFLRTADLPVWLAYKVAAAGISYH
DRKWCFCFGPR TNTILEDNNEVEVITKLGERKILRPRWADARVYSDHQAL
KSFKDFASGKRSQIGLVEV LGRMPEHFMGKTWEALDTMYVVATAEKGGRA
HMALEELPDALQTIALLSVM SLGVFFLLMQRKIGIKIGLGGVILGAATFFC
WMADVPGTKIAGMLLLSLLMIVLIP EPEKQRSQTDNQLAVFLICVLT
LVSAAVANEWGWDKTKNDISSLLGHKPEARETTLGVESF LLDLRPATAW
SLYAVTTAVLTPLLKHLITSYINTSLTSINVQASALFTLARGFPFVD
VGVSALLLAAGCWGQVTLTVTVTAAALLFCHYAYMVPGWQAEAMRSAQR
RTAAGIMKN AVVDGIVATDVPELERTTPVMQKKVGQIMLILVSMAAVVN
PNSVTVREAGILTTAAA VTLWENGASSVWNATTAIGLCHIMRGWLSCL
SITWTLIKMEKPGPKRGGA KGRTLGEVWKERLNMHTKEEFTRYRKEAITE
VDRSAAKHARREGNITGGHPVSRGTAKLRWLVE RRFLEPVGKVVDLGC
GRGGWCYMATQKRVQEVKGYTKGGPGHEEPQLVQSYGWNIVT MKSGVD
VFYRPEASDTLLCDIGESSSSAEVEEHRTVRVLEMVEDWLHRGPKEFCIK
V LCPYMPKVI EKMETLQRRYGGGLVRNPLSRNSTHEMYVWSHASGNIV
HSVNMTSQVLL GRMEKKTWKGPQFEEDVNLGSGTRAVGKPLLNSDT
SKIKNRIERLKKEYSSTWHQDAN HPYRTWNYHGSYEVKPTGSASSLVN
GVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVF KEKVDTKAPEPPEGV
KYVLNETTNWLWAF LARDKKPRMCSREEFIGKVN SNAALGAMFEEQNQ
WKNAREAVEDPKFWEMVDEERE AHLRGE CNTCIYNM MGKREKKPGEF
GKAKGS RAIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQ
KLGYLKEVGTKPGGK IYADDTAGWDTRITKADLENEAKVLELLDGEH
RLARSIIELTYRHKVVKVMRPAADG KTVMDVISREDQRGSGQVVTYAL
NTFTNLAVQLVRMMEGEGVIGPDDVEKL GKKGPK VRTWLFENGEE
RLSRMAVSGDDCVVKPLDDR FATSLHFLNAMS KVRKDIQEWKPS
TGW YDWQQVPFCSNHFTELIMKDGR TLVVP CRGQDELIGRARISPGAG
WNVRDTACLAKSY AQMWLLLYFHRRDLRLMANAICSAVPVNWVPTGR
TTWSIHAKGEWMTTEDMLS VWNRV WIEENEWMEDKTPVERWSDVPY
SGKREDIWCGLIGTRTRATWAENIHVAINQVRSVI GEEKYVDYMSSSL
RRYEDTIVVEDTVL"

mat_peptide	97. .411
	/product="C protein"
mat_peptide	412. .966
	/product="prM protein"
sig_peptide	412. .465
	/note="prM signal peptide"
mat_peptide	466. .741
	/product="cleaved amino terminal prM fragment"
mat_peptide	742. .966
	/product="M protein"
sig_peptide	964. .2472
	/note="E signal peptide"
mat_peptide	967. .2469

```

mat_peptide      /product="E protein"
                  2470. .3525
mat_peptide      /product="NS1 protein"
                  3526. .4218
mat_peptide      /product="NS2A protein"
                  4219. .4611
mat_peptide      /product="NS2B protein"
                  4612. .6468
mat_peptide      /product="NS3 protein"
                  6469. .6915
sig_peptide      /product="NS4A protein"
                  6847. .6915
mat_peptide      /note="2K peptide"
                  6916. .7683
mat_peptide      /product="NS4B protein"
                  7684. .10398
misc_feature     /product="NS5 protein"
                  2398. .2469
3'UTR           /note="NS1 signal peptide"
                  10402. .11048
ORIGIN

```

```

Query Match      87.4%;   Score 1327.8;   DB 10;   Length 11048;
Best Local Similarity 92.3%;   Pred: No. 0;
Matches 1398;   Conservative 0;   Mismatches 117;   Indels 0;   Gaps 0;

```

```

Qy      1 CGGAATTCAGCTTCAACTGTTTAGGAATGAGCAACAGGGACTTCCTGGAGGGAGTGTCTG 60
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      956 CAGCATACAGCTTCAACTGCTTAGGAATGAGCAACAGAGACTTCCTGGAGGGAGTGTCTG 1015

Qy      61 GAGCTACATGGGTTGATCTGGTACTGGAAGGAGACAGTTGTGTGACCATAATGTCAAAG 120
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1016 GAGCCACATGGGTTGATCTGGTACTGGAAGGCGACAGCTGTGTAACCATAATGTCAAAG 1075

Qy      121 ACAAGCCAACCATTGATGTCAAAATGATGAACATGGAAGCAGCTAATCTCGCAGATGTGC 180
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1076 ACAAGCCAACCATTGATGTCAAAATGATGAACATGGAAGCAGCTAATCTTGCAGATGTGC 1135

Qy      181 GTAGCTACTGCTACTTAGCTTCGGTCAGTGATCTGTCAACAAAAGCCGCGTGTCCAACCA 240
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1136 GCAGTTACTGTTACCTAGCTTCAGTCAGTGACTTGTCAACAAGAGCCGCGTGTCCAACCA 1195

Qy      241 TGGGTGAAGCTCACAACGAGAAAAGAGCCGACCCCTGCCTTTGTTTGCAAGCAAGGCGTCG 300
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1196 TGGGTGAAGCCCACAATGAAAAAAGAGCTGATCCCGCCTTCGTTTGCAAGCAAGGCGTTG 1255

Qy      301 TAGACAGAGGATGGGGGAATGGATGCGGACTGTTTGGAAAGGGGAGCATTGACACATGTG 360
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1256 TAGATAGAGGATGGGGAAACGGATGCGGACTGTTTGGAAAGGGGAGCATTGACACATGTG 1315

Qy      361 CAAAGTTTGCCTGTACAACCAAGGCAACTGGTTGGATTATCCAGAAGGAAAACATCAAGT 420
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1316 CGAAGTTTGCCTGTACAACCAAGCGACTGGTTGGATCATCCAGAAGGAAAACATCAAGT 1375

Qy      421 ACGAGGTTGCCATATTTGTGCATGGCCCCGACGACTGTGAATCACATGGCAATTATTCAA 480
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1376 ATGAGGTTGCCATATTTGTGCATGGCCCCGACGACCGTTGAATCTCATGGTGATTATTCAA 1435

```

Qy	481	CACAGATAGGGGCTACCCAAGCAGGAAGGTTGAGCATAACTCCATCGGCACCATCCTACA	540
Db	1436	CACAGATAGGGGCCACCCAGGCTGGAAGATTGAGCATAACTCCATCGGCGCCATCTTACA	1495
Qy	541	CGCTGAAGTTGGGTGAGTATGGTGAGGTCACAGTTGACTGTGAGCCACGGTCAGGAATAG	600
Db	1496	CGCTAAAGTTGGGTGAGTATGGTGAGGTAACGGTTGATTGTGAGCCACGGTCAGGAATAG	1555
Qy	601	ACACTAGCGCTTACTACGTTATGTCAGTGGGTGCGAAGTCCTTCTTGTTTACCCGAGAAT	660
Db	1556	ACACTAGCGCCTATTACGTTATGTCAGTTGGTGCGAAGTCCTTCTTGTTTACCCGAGAAT	1615
Qy	661	GGTTTATGGACCTGAACCTTCCATGGAGTAGCGCTGGAAGCACAACGTGGAGGAACCGGG	720
Db	1616	GGTTCATGGATCTGAACCTGCCATGGAGCAGTGCTGGAAGCACCACGTGGAGGAATCGGG	1675
Qy	721	AAACACTGATGGAGTTTGAAGAACCTCATGCCACCAAACAATCTGTCTAGCTCTAGGGT	780
Db	1676	AAACACTGATGGAGTTTGAAGAACCTCATGCCACCAAACGATCTGTTGTGGCTCTAGGGT	1735
Qy	781	CGCAGGAAGGTGCCTTGACCAAGCTCTGGCTGGAGCAATTCCTGTTGAGTTCTCAAGCA	840
Db	1736	CGCAGGAAGGCGCTTTGACCAAGCTCTGGCCGGAGCGATTCTGTTGAATTCTCAAGCA	1795
Qy	841	ACACTGTGAAGTTGACATCAGGACATCTGAAGTGTAGGGTGAAGATGGAGAAGTTGCAGC	900
Db	1796	ACACTGTGAAGTTGACATCAGGACATCTGAAGTGTAGGGTGAAGATGGAGAAGTTGCAGC	1855
Qy	901	TGAAGGGAACAACATATGGTGTATGCTCAAAAGCATTCAAATTCGCTAGGACTCCCCTG	960
Db	1856	TGAAGGGAACAACATACGGAGTATGTTCAAAAGCGTTCAAATTCGCTGGGACTCCTGCTG	1915
Qy	961	ACACTGGTCATGGAACGGTGGTGCTGGAAGTGCAGTATACCGGAAAAGACGGGCCTTGCA	1020
Db	1916	ACACTGGCCATGGAACGGTGGTGTTGGAAGTGCAGTACACCGGAACGGACGGTCCCTGCA	1975
Qy	1021	AAGTGCCCATTTCTTCTGTGGCTTCCCTGAACGACCTTACACCCGTTGGAAGGCTGGTGA	1080
Db	1976	AAGTGCCCATCTCTTCCGTAGCTTCCCTGAATGACCTCACACCTGTTGGAAGACTGGTAA	2035
Qy	1081	CTGTGAATCCATTTGTGTCTGTGGCTACGGCCAACCTCGAAGGTTTTGATTGAACTCGAAC	1140
Db	2036	CAGTGAATCCATTTGTGTCTGTGGCCACGGCCAACCTCGAAGGTTTTGATTGAACTCGAAC	2095
Qy	1141	CCCCGTTTAGTGACTCTTACATCGTGGTGGGGAGAGGAGAACAGCAGATAAACCACCACT	1200
Db	2096	CCCCGTTTCGGTGACTCCTACATCGTGGTGGGAAGAGGAGAACAGCAGATAAACCATCACT	2155
Qy	1201	GGCACAAATCTGGGAGCAGTATTGGAAAGGCTTTTACCACTACACTCAGAGGAGCTCAAC	1260
Db	2156	GGCACAAATCCGGGAGCAGCATTGGAAAGGCTTTTACTACCACACTCAGAGGAGCTCAAC	2215
Qy	1261	GACTTGCAGCTCTTGGAGACACTGCCTGGGATTTTGGATCAGTCGGAGGGGTTTTACCT	1320
Db	2216	GACTTGCAGCTCTTGGAGACACTGCTTGGGATTTTGGATCAGTTGGAGGGGTATTACCT	2275

```

Qy      1321  CGGTAGGGGAAAGCCATACACCAAGTTTTTGGAGGAGCCTTTAGATCACTCTTTGGAGGGA 1380
        |||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      2276  CGGTGGGGAAAGCTATACACCAAGTCTTTGGAGGAGCTTTTAGATCACTTTTGGAGGGA 2335

Qy      1381  TGTCTTGGATCACACAGGGGCTTCTGGGAGCTCTTCTGCTGTGGATGGGAATTAACGCCC 1440
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      2336  TGTCTTGGATCACACAGGGACTTCTGGGAGCTCTTCTGTTGTGGATGGGAATCAATGCCC 2395

Qy      1441  GTGACAGGTCAATTGCTATGACGTTTCCTTGCGGTTGGAGGAGTCTTGCTCTTCCTTTTCGG 1500
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      2396  GTGACAGGTCAATTGCTATGACGTTTCCTTGCGGTTGGAGGAGTTTGTCTCTTCCTTTTCGG 2455

Qy      1501  TCAACGTCCATGCTG 1515
        ||||| ||||
Db      2456  TCAACGTCCACGCTG 2470

```

RESULT 4

AY532665

LOCUS AY532665 11038 bp RNA linear VRL 09-DEC-2004

DEFINITION West Nile virus strain B956 polyprotein gene, complete genome.

ACCESSION AY532665

VERSION AY532665.1 GI:56462533

KEYWORDS .

SOURCE West Nile virus (WNV)

ORGANISM West Nile virus

Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Flavivirus; Japanese encephalitis virus group.

REFERENCE 1 (bases 1 to 11038)

AUTHORS Yamshchikov,G., Borisevich,V., Seregin,A., Chaporgina,E., Mishina,M., Mishin,V., Wai Kwok,C. and Yamshchikov,V.

TITLE An attenuated West Nile prototype virus is highly immunogenic and protects against the deadly NY99 strain: a candidate for live WN vaccine development

JOURNAL Virology 330 (1), 304-312 (2004)

PUBMED 15527855

REFERENCE 2 (bases 1 to 11038)

AUTHORS Borisevich,V.G. and Yamshchikov,V.F.

TITLE Molecular basis of attenuation of the West Nile virus prototype strain B956

JOURNAL Unpublished

REFERENCE 3 (bases 1 to 11038)

AUTHORS Borisevich,V.G. and Yamshchikov,V.F.

TITLE Direct Submission

JOURNAL Submitted (23-JAN-2004) Molecular Biosciences, University of Kansas, 1200 Sunnyside Ave., Lawrence, KS 66045, USA

FEATURES Location/Qualifiers

source 1. .11038

/organism="West Nile virus"

/mol_type="genomic RNA"

/strain="B956"

/db_xref="taxon:11082"

/note="obtained from R. Shope, Galveston, TX at suckling mouse brain passage 2; passaged once in C6/36 cells"

CDS 97. .10389

/codon_start=1

/product="polyprotein"

/protein_id="AAT02759.1"

/db_xref="GI:56462534"

/translation="MSKKPGGPGKNRAVNMLKRGMPRGLSLIGLKRAMLSLIDGKGPI
RFVLALLAFFRFTAIAPTRAVLDRWRGVNKTAMKHLISFKKELGTLTSAINRRSTKQ
KKRGGTAGFTILLGLIACAGAVTLSNFQGVMMTVNATDVTDVITPTAAGKNLCIVR
AMDVGYLCEDTITYECPVLAAGNDPEDIDCWCTKSSVYVRYGRCTKTRHSRRSRRSLT
VQTHGESTLANKKGAWL DSTKATRYLVKTESWILRNPYALVAAVIGWMLGSNTMQRV
VFAILLLLVPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTILMSKDKPTIDVK
MMNMEAANLADVRSYCYLASVSDLSTRAACPTMGEAHNEKRADPAFVCKQGVVDRGWG
NGCGLFGKGSIDTCAKFACTTKATGWIIQKENIKYEVAIFVHGPTTVESHGKIGATQA
GRFSITPSAPSYTLKLGEYGEVTVDCPEPRSGIDTSAYYVMSVGAKSFLVHREWFMDLN
LPWSSAGSTTWRNRETLMFEFEHPHATKQSVVALGSQEGALHQAALAGAIPEVEFSNTVK
LTSGLHKCRVKMEKLQKGTTYGVCSKAFKFARTPADTGHGTVVLELQYTGTDGPKV
PISSVASLNDLTPVGRLVTNPFVSVATANSKVLIELEPPFGDSYIVVGRGEQQINHH
WHKSGSSIGKAFTTTLRGAQRLAALGDTAWDFGSGGVFTSVGKAIHQVFGGAFRSLF
GGMSWITQGLLGALLLWMGINARDRSIAMTFLAVGGVLLFLSVNVHADTGCAIDIGRQ
ELRCGSGVFIHNDVEAWMDRYKFYPETPQGLAKIIQKAHAEGVCGLRSVSRLEHQMWE
AIKDELNTLLKENGVDLSVVVEKQNGMYKAAPKRLAATTEKLEMGWKAWGKSIIFAPE
LANNTFVIDGPETEECTANRAWNSMEVEDFGFGLTSTRMFLRIRETNTTECDSKIIG
TAVKNNMAVHSDLSYWIESGLNDTWKLERAVLGEVKSCTWPETHTLWGDGVLESLLII
PITLAGPRSNHNRRPGYKTQNGQGPWDEGRVEIDFDYCPGTTVTISDSCGHRGPAARTT
TESGKLITDWCCRSCTLPPLRFQTENGCWYGMIEIRPTRHDEKTLVQSRVNAYNADMID
PFQLGLLVVFLATQEVLRKRWTAKISIPAIMLALLVLVFGGITYTDVLRVILVGAFA
AEANSGGDVVHLALMATFKIQPVFLVASFLKARWTNQESILLMLAAFFQMAYYDAKN
VLSWEVPDVLNLSVAWMILRAISFTNTSNVVVPLLALLTPGLKCLNLDVYRILLMV
GVGSLIKEKRSSAAKKKGACLICLALASTGVFNPMILAAGLMACDPNRKRGPATEVM
TAVGLMFAIVGGLAELDIDSMAPMTIAGLMFVAFVISGKSTDMWERTADITWESDA
EITGSSSERVDVRLDDDGNFQLMNDPGAPWKIWMMLRMACLAISAYTPWAILPSVIGFWI
TLQYTKRGGVLWDTSPSPKEYKKGDTTGTGYRIMTRGLLGSYQAGAGVMVEGVFHTLWH
TTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGHDEVQMI VVEPGKNVKNV
QTKPGVFKTPEGEIGAVTLDYPTGTSGSPIVDKNGDVI GLYGNVIMPNGSYISAI VQ
GERMEEPAPAGFEPEMLRKKQITVLDLHPGAGKTRKILPQIIKEAINKRLRTAVLAPT
RVVAAEMSEALRGLPIRYQTSAVHREHSGNEIVDMCHATLTHRLMSPHRVPNYNLFIT
MDEAHFTDPASIAARGYIATKVELGEAAAI FMTATPPGTSDFPFESNAPISDMQTEIP
DRAWNTGYEWITEYVGKTVWFVPSVKMGNEIALCLQRAGKKVIQLNRKSYETEPKCK
NDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIIIEEGDGRVILGEPSAITAASAAQ
RRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLVAQLYQPERE
KVYTMDEYRLRGEERKNFLEFLRTADLPVWLAYKVAAAGISYHDRKWCDFGPRNTTI
LEDNNEVEVITKLGERKILRPRWADARVYSDHQALKSFKDFASGKRSQIGLVEVLGRM
PEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALLSVMISLGVFFL
LMQRKGIKIGLGGIILGAATFFCWMAEVPGTKIAGMLLLSLLMIVLIPPEPEKQRSQ
TDNQLAVFLICVLTLVGAVANEMGWLDKTKNDISSLLGHKPEARETTLGVESFLLDL
RPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALEFTLARGFPFVDVGVS
ALLLAAGCWGQVTLTVTVTAAALLFCHYAYMVPGWQAEAMRSAQRRRTAAGIMKNVVVD
GIVATDVPELERTTPVMQKKVGQIILILVSMAAVVNPSVRTVREAGILTTAAAVTLW
ENGASSVWNATTAIGLCHIMRGGWLSCLSIMWTLIKNMEKPGLKRGGAKGRTLGEVWK
ERLNHMTKEEFTRYRKEAITEVDRSAAKHARREGNITGGHPVSRGTAKLRWLVERRFL
EPVGKVVDLGCGRGGWCYMATQKRVQEVKGYTKGGPGHEEPQLVQSYGWNIVTMKSG
VDVFYRPSEASDTLLCDIGESSSSAEVEEHRTVRVLEMVEDWLHRGPKEFCIKVLCPY
MPKVI EKMETLQRRYGGGLVRNPLSRNSTHEMYWVSHASGNIVHSVNMTSQVLLGRME
KKTWKGPQFEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLKKEYSSSTWHQDANHPYR
TWNHYHGSYEVKPTGSASSLVNGVVRLLSKPWDITITNVTTMAMTDTTFPGQQRVFEKEV
DTKAPEPPEGVKYVLNETTNWLWAF LARDKKPRMCSREEFIGKVSNAALGAMFEEQN
QWKNAREAVEDPKFWEMVDEEREHLRGECENTCIYNNMGKREKKPGFEFGKAKGSRATW
FMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQLKGYILKEVGTKPGGKVYAD
DTAGWDTRITKADLENEAKVLELLDGEHRLARSIIELTYRHKKVVKMRPAADGKTVM
DVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLGKKGPKVVRTW

Db	1604	GGTTTATGGATCTGAACCTGCCATGGAGCAGTGCTGGAAGCACACGCTGGAGGAACCGGG	1663
Qy	721	AAACACTGATGGAGTTTGAAGAACCTCATGCCACCAAACAATCTGTCTAGCTCTAGGGT	780
Db	1664	AAACACTGATGGAGTTTGAAGAACCTCATGCCACCAAACAATCTGTTGTGGCTCTAGGGT	1723
Qy	781	CGCAGGAAGGTGCCTTGCACCAAGCTCTGGCTGGAGCAATTCCTGTTGAGTTCTCAAGCA	840
Db	1724	CGCAGGAAGGTGCGTTGCACCAAGCTCTGGCCGGAGCGATTCTGTTGAGTTCTCAAGCA	1783
Qy	841	ACACTGTGAAGTTGACATCAGGACATCTGAAGTGTAGGGTGAAGATGGAGAAGTTGCAGC	900
Db	1784	ACACTGTGAAGTTGACATCAGGACATCTGAAGTGTAGGGTGAAGATGGAGAAGTTGCAGC	1843
Qy	901	TGAAGGGAACAACATATGGTGTATGCTCAAAAGCATTCAAATTCGCTAGGACTCCCGCTG	960
Db	1844	TGAAGGGAACAACATATGGAGTATGTTCAAAGCGTTCAAATTCGCTAGGACTCCCGCTG	1903
Qy	961	ACACTGGTCATGGAACGGTGGTGTCTGGAAGTGCAGTATACCGGAAAAGACGGGCCTTGCA	1020
Db	1904	ACACTGGCCACGGAACGGTGGTGTGGAAGTGAATATACCGGAACAGACGGTCCCTGCA	1963
Qy	1021	AAGTGCCCATTTCTTCTGTGGCTTCCCTGAACGACCTTACACCCGTTGGAAGGCTGGTGA	1080
Db	1964	AAGTGCCCATTTCTTCCGTAGCTTCCCTGAATGACCTCACACCTGTTGGAAGACTGGTGA	2023
Qy	1081	CTGTGAATCCATTTGTGTCTGTGGCTACGGCCAACTCGAAGGTTTTGATTGAACTCGAAC	1140
Db	2024	CCGTGAATCCATTTGTGTCTGTGGCCACAGCCAACTCGAAGGTTTTGATTGAACTCGAAC	2083
Qy	1141	CCCCGTTTGTAGTACTCTTACATCGTGGTGGGGAGAGGAGAACAGCAGATAAACCACCACT	1200
Db	2084	CCCCGTTTGGTGACTCTTACATCGTGGTGGGAAGAGGAGAACAGCAGATAAACCATCACT	2143
Qy	1201	GGCACAAATCTGGGAGCAGTATTGGAAAGGCTTTACCACTACACTCAGAGGAGCTCAAC	1260
Db	2144	GGCACAAATCTGGGAGCAGCATTGGAAAGGCTTTACCACCACACTCAGAGGAGCTCAAC	2203
Qy	1261	GACTTGCAGCTCTTGGAGACACTGCCTGGGATTTTGGATCAGTCGGAGGGGTTTTACCT	1320
Db	2204	GACTCGCAGCTCTTGGAGATACTGCTTGGGATTTTGGATCAGTTGGAGGGGTTTTACCT	2263
Qy	1321	CGGTAGGGAAAGCCATACACCAAGTTTTTGGAGGAGCCTTTAGATCACTCTTTGGAGGGA	1380
Db	2264	CAGTGGGGAAAGCCATACACCAAGTCTTTGGAGGAGCTTTTAGATCACTCTTTGGAGGGA	2323
Qy	1381	TGTCCTGGATCACACAGGGGCTTCTGGGAGCTCTTCTGCTGTGGATGGGAATTAACGCCC	1440
Db	2324	TGTCCTGGATCACACAGGGACTTCTGGGAGCTCTTCTGTTGTGGATGGGAATCAATGCCC	2383
Qy	1441	GTGACAGGTCAATTGCTATGACGTTTCTTGCAGTTGGAGGAGTCTTGCTCTTCTTTTCGG	1500
Db	2384	GTGACAGGTCAATTGCTATGACGTTTCTTGCAGTTGGAGGAGTTTTGCTCTTCTTTTCGG	2443
Qy	1501	TCAACGTCCATGCTG	1515
Db	2444	TCAACGTCCATGCTG	2458

RESULT 5

DQ318019

LOCUS DQ318019 11038 bp mRNA linear VRL 01-JAN-2006

DEFINITION West Nile virus strain ArD76104, complete genome.

ACCESSION DQ318019

VERSION DQ318019.1 GI:84028432

KEYWORDS .

SOURCE West Nile virus (WNV)

ORGANISM West Nile virus

Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Flavivirus; Japanese encephalitis virus group.

REFERENCE 1 (bases 1 to 11038)

AUTHORS Borisevich,V.G., Seregin,A.V. and Yamshchikov,V.F.

TITLE Genetic determinants of West Nile virus pathogenicity

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 11038)

AUTHORS Borisevich,V.G. and Yamshchikov,V.F.

TITLE Direct Submission

JOURNAL Submitted (07-DEC-2005) Molecular Biosciences, 1200 Sunnyside ave, Lawrence, KS 66045, USA

FEATURES Location/Qualifiers

source

1. .11038

/organism="West Nile virus"

/mol_type="mRNA"

/strain="ArD76104"

/db_xref="taxon:11082"

/country="Senegal"

/note="lineage 2; SMB pass 3, C6/36 pass 1"

5'UTR

1. .96

CDS

97. .10389

/codon_start=1

/product="polyprotein"

/protein_id="ABC49716.1"

/db_xref="GI:84028433"

/translation="MSKKPGGPGKNRAVNMLKRGMPRGLSLIGLKRAMLSLIDGKGPI
RFVLALLAFFRFTAIAPTRAVLDRWRGVNKTAMKHLISFKKELGTLTSAINRRSTKQ
KKRGGTAGFTILLGLIACAGAVTLSNFQGVMMTVNATDVTDVITPTAAGKNLCIVR
AMDVGYLCEDTITYECPVLAAGNDPEDIDCWCTKSSVYVRYGRCTKTRHSRRSRRSLT
VQTHGESTLANKKGAWLDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
VFAILLLLIVAPAYSFNCGLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
MMNMEAANLADVRSYCYLASVSDLSTRAACPTMGEAHNEKRADPAFVCKQGVVDRGWG
NGCGLFGKGSIDTCAKFACTTKATGWIIQKENIKYEVAIFVHGPTTVESHGKIGATQA
GRFSITPSAPSYTLKLGEYGEVTVDCPRSGIDTSAYYVMSVGAKSFLVHREWFMDLN
LPWSSAGSTTWRNRETLVEFEPPHATKQSVVALGSQEGALHQAALAGAI PVEFSSNTVK
LTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFARTPADTGHGTVVLELQYTGTGDPCKV
PISSVASLNDLTPVGRVLTVPFVSVATANSKVLIELEPPFGDSYIVVGRGEQQINHH
WHKSGSSIGKAFTTTLRGAQRLAALGDTAWDFGSGGVFTSVGKAIHQVFGGAFRSLF
GGMSWITQGLLGALLLWMGINARDRSIAMTFLAVGGVLLFLSVNVHADTGCAIDIGRQ
ELRCGSGVFIHNDVEAWMDRYKFYPETPQGLAKIIQKAHAEGVCGLRSVSRLEHQMWE
AIKDELNTLLKENGVDLSVVVEKQNGMYKAAPKRLAATTEKLEMGWKAWSKIIFAPE
LANNTFVIDGPETEECTANRAWNSMEVEDFGFGLTSTRMFLIRRETNTTECDSKIIG
TAVKNNMAVHSDLSYWIESGLNDTWKLERAVLGEVKSCTWPETHTLWGDGVLESDLII
PITLAGPRSNHNRPGYKTQNGQPWDEGRVEIDFDYCPGTTVTISDSCGHRGPAARTT
TESGKLITDWCCRSTLPPLRFQTEGCWYGMEIRPTRHDEKTLVQSRVNAYNADMID
PFQLGLLVVFLATQEVLRKRWTAKISIPAIMLALLVLFVGGITYTDVLRVYILVGAAF

AEANSGGDVVHLALMATFKIQPVFLVASFLKARWTNQESILLMLAAAFFQMAYYDAKN
VLSWEVPDVLSLSVAWMILRAISFTNTSNVVPPLALLTPGLKCLNLDVYRIILLMV
GVGSLIKEKRSSAAKKKGACLICLALASTGVFNPMILAAGLMACDPNRKRGWPATEVM
TAVGLMFAIVGGLAELDIDSMAPMTIAGLMFVAFVISGKSTDMWIERTADITWESDA
EITGSSSERVDVRLDDDGNFQLMNDPGAPWKIWMLRMACLAISAYTPWAILPSVIGFWI
TLQYTKRGGVLWDTPSPKEYKKGDTTGTGVYRIMTRGLLGSYQAGAGVMVEGVFHTLWH
TTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGHDEVQMIVVEPGKNVKNV
QTKPGVFKTPEGEIGAVTLDYPTGTSGSPIVDKNGDVI GLYNGVIMPNGSYISAIVQ
GERMEEPAPAGFEPEMLRKKQITVLDLHPGAGKTRKILPQIIKEAINKRLRTAVLAPT
RVVAAEMSEALRGLPIRYQTSAVHREHSGNEIVDMCHATLTHRLMSPHRVPNYNLFI
MDEAHFTDPASIAARGYIATKVELGEAAAI FMTATPPGTSDFPFESNAPISDMQTEIP
DRAWNTGYEWITEYVGKTVWFVPSVKMGNEIALCLQRAGKKVIQLNRKSYETEYPKCK
NDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIIIEEGDGRVILGEPSAITAASAAQ
RRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLVAQLYQPERE
KVYTMDEYRLRGEERKNFLEFLRTADLPVWLAYKVAAAGISYHDRKWCDFGPRNTTI
LEDNNEVEVITKLGERKILRPWADARVYSDHQALKSFKDFASGKRSQIGLVEVLGRM
PEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALLSVM SLGVFFL
LMQRKGIGKIGLGGVILGAATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPPEKQRSQ
TDNQLAVFLICVLTLSAVAANEMGWLDKTKNDIGSLLGHKPEARETTLGVESFLLDL
RPATAWSLYAVTTAVLTPLLKHLITSYINTSLTSINVQASALFTLARGFPFVDVGV
ALLLAAGCWGQVTLTVTVTAAALLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNAVVD
GIVATDVPELERTTPVMQKKVGQIMLILVSMAAVVNPSVRTVREAGILTAAAVTLW
ENGASSVWNATTAIGLCHIMRGGWLSCLSIMWTLIKMEKPGPKRGGAKGRTLGEVWK
ERLNHMTKEEFTRYRKEAITEVDRSAKHARREGNITGGHPVSRGTAKLRWLVERRFL
EPVGKVVDLGCGRGGWCYYMATQKRVQEVKGYTKGGPGHEEPQLVQSYGWNIVTMKSG
VDVFYRPSEASDTLLCDIGESSSSAEVEEHRTVRVLEMVEDWLHRGPKEFCIKVLCPY
MPKVIEKMETLQRRYGGGLVRNPLSRNSTHEMYWVSHASGNIVHSVNMTSQVLLGRME
KKTWKGPQFEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLKKEYSSTWHQDANHPYR
TWNHYGSYEVKPTGSASSLVNGVVRLLSKPWDITITNVTTMAMTDTTPFGQQRVFKEKV
DTKAPEPPEGVKYVLNETTNWLWAF LARDKKPRMCSREEFIGKVN SNAALGAMFEEQN
QWKNAREAVEDPKFWEMVDEEREHLRGECENTCIYNMMGKREKKPGFEFGKAKGSRAIW
FMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILKEVGT KPGGKVYAD
DTAGWDTRITKADLENEAKVLELLDGEHRR LARSIIELTYRHKVVKMRPAADGKTVM
DVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLKGKGPVKVRTW
LFENGEERLSRMAVSGDDCVKPLDDR FATSLHFLNAMS KVRKDIQEWKPSTGWYDWQ
QVPFCSNHFTELIMKDGRTLVVP CRGQDELIGRARISPGAGWNVRDTACLAKSYAQMW
LLLYFHRRDLRLMANAICSAVPVNWVPTGR TTWSIHAKGEWMTTEDMLAVWNRVWIEE
NEWMEDKTPVERWSDVPYSGKREDIWC GSLIGTRTRATWAENIHVAINQVRSVIGEEK
YVDYMSSLLRRYEDTIVVEDTVL"

mat_peptide	97. .411
	/product="C protein"
sig_peptide	412. .465
	/note="prM signal peptide"
mat_peptide	466. .966
	/product="prM protein"
mat_peptide	466. .741
	/product="cleaved amino terminal prM fragment"
mat_peptide	742. .966
	/product="M protein"
sig_peptide	919. .966
	/note="E signal peptide"
mat_peptide	967. .2457
	/product="E protein"
sig_peptide	2386. .2457
	/note="NS1 signal peptide"
mat_peptide	2458. .3513
	/product="NS1 protein"

```

mat_peptide      3514. .4206
                  /product="NS2A protein"
mat_peptide      4207. .4599
                  /product="NS2B protein"
mat_peptide      4600. .6456
                  /product="NS3 protein"
mat_peptide      6457. .6903
                  /product="NS4A protein"
sig_peptide      6835. .6903
                  /note="2K peptide"
mat_peptide      6904. .7671
                  /product="NS4B protein"
mat_peptide      7672. .10386
                  /product="NS5 protein"
3'UTR            10390. .11038

```

ORIGIN

```

Query Match      86.9%;  Score 1321;  DB 10;  Length 11038;
Best Local Similarity 92.6%;  Pred. No. 0;
Matches 1403;  Conservative 0;  Mismatches 100;  Indels 12;  Gaps 1;

```

```

Qy      1 CGGAATTCAGCTTCAACTGTTTAGGAATGAGCAACAGGGACTTCCTGGAGGGAGTGTCTG 60
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     956 CAGCATACAGCTTCAACTGCTTAGGAATGAGTAACAGAGACTTCCTGGAGGGAGTGTCTG 1015

Qy      61 GAGCTACATGGGTTGATCTGGTACTGGAAGGAGACAGTTGTGTGACCATAATGTCAAAAG 120
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    1016 GAGCTACATGGGTTGATCTGGTACTGGAAGGCGATAGTTGTGTGACCATAATGTCAAAAG 1075

Qy     121 ACAAGCCAACCATTGATGTCAAAATGATGAACATGGAAGCAGCTAATCTCGCAGATGTGC 180
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    1076 ACAAGCCAACCATTGATGTCAAAATGATGAACATGGAAGCAGCCAACCTCGCAGATGTGC 1135

Qy     181 GTAGCTACTGCTACTTAGCTTCGGTCAGTGATCTGTCAACAAAAGCCGCGTGTCCAACCA 240
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    1136 GCAGTTACTGTTACCTAGCTTCGGTCAGTGACTTGTCAACAAGAGCTGCGTGTCCAACCA 1195

Qy     241 TGGGTGAAGCTCACAACGAGAAAAGAGCCGACCCTGCCTTTGTTTGCAAGCAAGGCGTGC 300
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    1196 TGGGTGAAGCCCACAACGAGAAAAGAGCTGACCCCGCCTTCGTTTGCAAGCAAGGCGTTG 1255

Qy     301 TAGACAGAGGATGGGGGAATGGATGCGGACTGTTTGGAAGGGGAGCATTGACACATGTG 360
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    1256 TGGACAGAGGATGGGGAAATGGCTGCGGACTGTTTGGAAGGGGAGCATTGACACATGTG 1315

Qy     361 CAAAGTTTGCCTGTACAACCAAGGCAACTGGTGGATTATCCAGAAGGAAAACATCAAGT 420
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    1316 CGAAGTTTGCCTGTACAACCAAGGCAACTGGATGGATCATCCAGAAGGAAAACATCAAGT 1375

Qy     421 ACGAGGTTGCCATATTTGTGCATGGCCCCGACGACTGTGCAATCACATGGCAATTATTCAA 480
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    1376 ATGAGGTTGCCATATTTGTGCATGGCCCCGACGACCGTTGAATCTCATGGCA----- 1426

Qy     481 CACAGATAGGGGCTACCCAAGCAGGAAGGTTTCAGCATAACTCCATCGGGCACCATCCTACA 540
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    1427 ---AGATAGGGGGCCACCCAGGCTGGAAGATTTCAGTATAACTCCATCGGGCGCCATCTTACA 1483

```

Qy	541	CGCTGAAGTTGGGTGAGTATGGTGAGGTCACAGTTGACTGTGAGCCACGGTCAGGAATAG	600
Db	1484	CGCTAAAGTTGGGTGAGTATGGTGAGGTTACGGTTGATTGTGAGCCACGGTCAGGAATAG	1543
Qy	601	ACACTAGCGCTTACTACGTTATGTCAGTGGGTGCGAAGTCCTTCTTGGTTCACCGAGAAT	660
Db	1544	ACACTAGCGCCTATTACGTTATGTCAGTTGGTGCGAAGTCCTTCTTGGTTCACCGAGAAT	1603
Qy	661	GGTTTATGGACCTGAACCTTCCATGGAGTAGCGCTGGAAGCACAAACGTGGAGGAACCGGG	720
Db	1604	GGTTTATGGATCTGAACCTGCCATGGAGCAGTGCTGGAAGCACACGTGGAGGAACCGGG	1663
Qy	721	AAACACTGATGGAGTTTGAAGAACCTCATGCCACCAAACAATCTGTCTAGCTCTAGGGT	780
Db	1664	AAACACTGGTGGAGTTTGAAGAACCTCATGCCACCAAACAATCTGTTGTGGCTCTAGGGT	1723
Qy	781	CGCAGGAAGGTGCCTTGCACCAAGCTCTGGCTGGAGCAATTCCTGTTGAGTTCTCAAGCA	840
Db	1724	CGCAGGAAGGTGCGTTGCACCAAGCTCTGGCCGGAGCGATTCTGTTGAGTTCTCAAGCA	1783
Qy	841	ACACTGTGAAGTTGACATCAGGACATCTGAAGTGTAGGGTGAAGATGGAGAAGTTGCAGC	900
Db	1784	ACACTGTGAAGTTGACATCAGGACATCTGAAGTGTAGGGTGAAGATGGAGAAGTTGCAGC	1843
Qy	901	TGAAGGGAACAACATATGGTGTATGCTCAAAAGCATTCAAATTCGCTAGGACTCCCGCTG	960
Db	1844	TGAAGGGAACAACATATGGAGTATGTTCAAAAGCGTTCAAATTCGCTAGGACTCCCGCTG	1903
Qy	961	ACACTGGTCATGGAACGGTGGTGCTGGAAGTGCAGTATACCGGAAAAGACGGGCCTTGCA	1020
Db	1904	ACACTGGCCACGGAACGGTGGTGCTGGAAGTGCAGTATACCGGAACAGACGGTCCCTGCA	1963
Qy	1021	AAGTGCCCATTTCTTCTGTGGCTTCCCTGAACGACCTTACACCCGTTGGAAGGCTGGTGA	1080
Db	1964	AAGTGCCCATTTCTTCCGTAGCTTCCCTGAATGACCTCACACCTGTTGGAAGACTGGTGA	2023
Qy	1081	CTGTGAATCCATTTGTGTCTGTGGCTACGGCCAACCTCGAAGGTTTTGATTGAACTCGAAC	1140
Db	2024	CCGTGAATCCATTTGTGTCTGTGGCCACAGCCAACCTCGAAGGTTTTGATTGAACTCGAAC	2083
Qy	1141	CCCCGTTTAGTGACTCTTACATCGTGGTGGGGAGAGGAGAACAGCAGATAAACCACCACT	1200
Db	2084	CCCCGTTTGGTGACTCTTACATCGTGGTGGGAAGAGGAGAACAGCAGATAAACCATCACT	2143
Qy	1201	GGCACAAATCTGGGAGCAGTATTGGAAGGCTTTACCACTACACTCAGAGGAGCTCAAC	1260
Db	2144	GGCACAAATCTGGGAGCAGCATTGGAAGGCTTTACCACTACACTCAGAGGAGCTCAAC	2203
Qy	1261	GACTTGCAGCTCTTGGAGACACTGCCTGGGATTTTGGATCAGTCGGAGGGGTTTTACCT	1320
Db	2204	GACTCGCAGCTCTTGGAGATACTGCTTGGGATTTTGGATCAGTTGGAGGGGTTTTACCT	2263
Qy	1321	CGGTAGGGAAAGCCATACACCAAGTTTTTGGAGGAGCCTTTAGATCACTCTTTGGAGGGA	1380
Db	2264	CAGTGGGGAAAGCCATACACCAAGTCTTTGGAGGAGCTTTTAGATCACTCTTTGGAGGGA	2323
Qy	1381	TGTCCTGGATCACACAGGGGCTTCTGGGAGCTCTTCTGCTGTGGATGGGAATTAACGCCC	1440

```

          |||
Db      2324 TGTCTGGATCACACAGGGACTTCTGGGAGCTCTTCTGTTGTGGATGGGAATCAATGCCC 2383
          |||
Qy      1441 GTGACAGGTCAATTGCTATGACGTTCTTGCAGTTGGAGGAGTCTTGCTCTTCCTTTTCGG 1500
          |||
Db      2384 GTGACAGGTCAATTGCTATGACGTTTCTTGCAGTTGGAGGAGTTTGTCTCTTCCTTTTCGG 2443
          |||
Qy      1501 TCAACGTCCATGCTG 1515
          |||
Db      2444 TCAACGTCCATGCTG 2458

```

RESULT 6

WNFCG

LOCUS WNFCG 10962 bp ss-RNA linear VRL 08-MAY-2002

DEFINITION West Nile virus RNA, complete genome.

ACCESSION M12294 M10103

VERSION M12294.2 GI:11497619

KEYWORDS .

SOURCE West Nile virus

ORGANISM West Nile virus

Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Flavivirus; Japanese encephalitis virus group.

REFERENCE 1 (bases 67 to 969)

AUTHORS Castle,E., Nowak,T., Leidner,U., Wengler,G. and Wengler,G.

TITLE Sequence analysis of the viral core protein and the membrane-associated proteins V1 and NV2 of the flavivirus West Nile virus and of the genome sequence for these proteins

JOURNAL Virology 145 (2), 227-236 (1985)

PUBMED 2992152

REFERENCE 2 (bases 859 to 2658)

AUTHORS Wengler,G., Castle,E., Leidner,U., Nowak,T. and Wengler,G.

TITLE Sequence analysis of the membrane protein V3 of the flavivirus West Nile virus and of its gene

JOURNAL Virology 147 (2), 264-274 (1985)

PUBMED 3855247

REFERENCE 3 (bases 1 to 10962)

AUTHORS Castle,E.

JOURNAL Unpublished

REFERENCE 4 (bases 67 to 10485)

AUTHORS Castle,E., Leidner,U., Nowak,T., Wengler,G. and Wengler,G.

TITLE Primary structure of the West Nile flavivirus genome region coding for all nonstructural proteins

JOURNAL Virology 149 (1), 10-26 (1986)

PUBMED 3753811

REFERENCE 5 (bases 1 to 10962)

AUTHORS Yamshchikov,V.F., Wengler,G., Perelygin,A.A., Brinton,M.A. and Compans,R.W.

TITLE An infectious clone of West Nile flavivirus

JOURNAL Virology (2000) In press

REFERENCE 6 (bases 1 to 10962)

AUTHORS Castle,E.

TITLE Direct Submission

JOURNAL Submitted (03-AUG-1993) Justus-Liebig-Universitat Giessen, Institut fur Virologie, 35392, Giessen, Germany

REFERENCE 7 (bases 1 to 10962)

AUTHORS Yamshchikov,V.F.

TITLE Direct Submission
 JOURNAL Submitted (01-DEC-2000) University of Virginia Health Sciences
 Centre, Department of Internal Medicine/GI, Charlottesville, VA
 22906
 COMMENT On Dec 1, 2000 this sequence version replaced gi:336167.
 Draft entry and sequence in computer readable form for
 [1],[2],[4],[3] kindly provided by E.Castle, 12-NOV-1985. The West
 Nile viral genome consists of a 42S viral RNA. The amino-terminal
 ends of the structural proteins were experimentally determined. An
 'atg' codon is located at positions 142-144, which could be used
 for an alternative initiation of translation for V2. The
 carboxy-terminal ends of the proteins reported here were not yet
 precisely defined.
 FEATURES Location/Qualifiers
 source 1. .10962
 /organism="West Nile virus"
 /virion
 /mol_type="genomic RNA"
 /db_xref="taxon:11082"
 /clone="33/G8; 34/F6"
 CDS 97. .10389
 /codon_start=1
 /product="polyprotein precursor"
 /protein_id="AAA48498.2"
 /db_xref="GI:11497620"
 /translation="MSKKPGGPGKNRAVNMLKRGMPRGLSLIGLKRAMLSLIDGKGPI
 RFVLALLAFFRFTAIPTRAVLDRWGVNKQTAMKHLLSFKKELGTLTSAINRRSTKQ
 KKRGGTAGFTILLGLIACAGAVTLSNFQGVMMTVNATDVTDVITPTAAGKNLCIVR
 AMDVGYLCEDTITYECPVLAAGNDPEDIDCWCTKSSVYVRYGRCTKTRHSRRSRRSLT
 VQTHGESTLANKKGAWLDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
 VFAILLLLAVAPYSFNCIGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
 MMNMEAANLADVRSYCYLASVSDLSTRAACPTMGEAHNEKRADPAFVCKQGVDVDRGWG
 NGCGLFGKGSIDTCAKFACTTKATGWIIQKENIKYEVAIFVHGPTTVESHGKIGATQA
 GRFSITPSAPSYTLKLGEYGEVTVDCPEPRSGIDTSAYYVMSVGEKSFLVHREWFMDLN
 LPWSSAGSTTWRNRETLMEFEHPHATKQSVVALGSQEGALHQALAGAI PVEFS SNTVK
 LTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFARTPADTGHGTVVLELQYTGTDGPCKV
 PISSVASLNDLTPVGRLVTNPFVSVATANSKVLIELEPPFGDSYIVVGRGEQQINHH
 WHKSGSSIGKAFTTTLRGAQRLAALGDTAWDFGSGGVFTSVGKAIHQVFVGGAFRSLF
 GGMSWITQGLLGLALLWMGINARDRSIAMTFLAVGGVLLFLSVNVHADTGCAIDIGRQ
 ELRCGSGVFIHNDVEAWMDRYKFYPETPQGLAKIIQKAHAEGVCGLRSVSRLEHQMWE
 AIKDELNTLLKENGVDLSVVVEKQNGMYKAAPKRLAATTEKLEMGWKAWGKSIIFAPE
 LANNTFVIDGPETEECTANRAWNSMEVEDFGFGLTSTRMFLRIRETNTTECDSKIIG
 TAVKNNMAVHSDLSYWIESGLNDTWKLERAVLGEVKSCTWPETHTLWGDGVLESDLII
 PITLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTISDSCEHRGPAARTT
 TESGKLITDWCCRSC TLPPLRFQTENG CWYGM EIRPTRHDEKTLVQSRVNAYNADMID
 PFQLGLMVVFLATQEVLRKRWTAKISIPAIMLALLVLVFGGITYTDVLRVILVGAAF
 AEANSGGDVVHLALMATFKIQPVFLVASFLKARWTNQESILLMLAAFFQMAYYDAKN
 VLSWEVPDVLNSLSVAWMILRAISFTNTSNVVVPLLALLTPGLKCLNLDVYRILLMV
 GVGS LIKEKRSSAAKKKGACLICLALASTGVFNPMILAAGLMACDPNRKRGPATEVM
 TAVGLMFAIVGGLAELDIDSMAPMTIAGLMFAAFVISGKSTDMWIERTADITWESDA
 EITGSSSERVDVRLDDGDNFQLMNDPGAPWKIWMLRMACLAISAYTPWAILPSVIGFWI
 TLQYTKRGGVLWDTPSPKEYKKGDTTGTGYRIMTRGLLGSYQAGAGVMVEGVFHTLWH
 TTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGHDEVQMI VVEPGKNVKNV
 QTKPGVFKTPEGEIGAVTLDYPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYISAI VQ
 GERMEEPAPAGFEP EMLRKKQITVLDLHPGAGKTRKILPQIIKEAINKRLRTAVLAPT
 RVVAAEMSEALRGLPIRYQTS AVHREHSGNEIVDMCHATLTHRLMSPHRVPNYNLF I
 MDEAHFTDPASIAARGYIATKVELGEAAAAI FMTATPPGTSDFPFESNAPISDMQTEIP

Qy	61	GAGCTACATGGGTTGATCTGGTACTGGAAGGAGACAGTTGTGTGACCATAATGTCAAAAAG	120
Db	1016	GAGCTACATGGGTTGATCTGGTACTGGAAGGCGATAGTTGTGTGACCATAATGTCAAAAAG	1075
Qy	121	ACAAGCCAACCATTTGATGTCAAAATGATGAACATGGAAGCAGCTAATCTCGCAGATGTGC	180
Db	1076	ACAAGCCAACCATTTGATGTCAAAATGATGAACATGGAAGCAGCCAACCTCGCAGATGTGC	1135
Qy	181	GTAGCTACTGCTACTTAGCTTCGGTCAGTGATCTGTCAACAAAAGCCGCGTGTCCAACCA	240
Db	1136	GCAGTTACTGTTACCTAGCTTCGGTCAGTGACTTGTCAACAAGAGCTGCGTGTCCAACCA	1195
Qy	241	TGGGTGAAGCTCACAAACGAGAAAAGAGCCGACCCTGCCTTTGTTTGCAAGCAAGGCGTCG	300
Db	1196	TGGGTGAAGCCCACAACGAGAAAAGAGCTGACCCCGCCTTCGTTTGCAAGCAAGGCGTTG	1255
Qy	301	TAGACAGAGGATGGGGGAATGGATGCGGACTGTTTGGAAGGGGAGCATTGACACATGTG	360
Db	1256	TGGACAGAGGATGGGGGAAATGGCTGCGGACTGTTTGGAAGGGGAGCATTGACACATGTG	1315
Qy	361	CAAAGTTTGCCTGTACAACCAAGGCAACTGGTTGGATTATCCAGAAGGAAAACATCAAGT	420
Db	1316	CGAAGTTTGCCTGTACAACCAAGCAACTGGATGGATCATCCAGAAGGAAAACATCAAGT	1375
Qy	421	ACGAGGTTGCCATATTTGTGCATGGCCCCGACGACTGTGCAATCACATGGCAATTATTCAA	480
Db	1376	ATGAGGTTGCCATATTTGTGCATGGCCCCGACGACCGTTGAATCTCATGGCA-----	1426
Qy	481	CACAGATAGGGGCTACCCAAGCAGGAAGGTTTACGCATAACTCCATCGGCACCATCCTACA	540
Db	1427	---AGATAGGGGGCCACCCAGGCTGGAAGATTACGTATAACTCCATCGGCGCCATCTTACA	1483
Qy	541	CGCTGAAGTTGGGTGAGTATGGTGAGGTCACAGTTGACTGTGAGCCACGGTCAGGAATAG	600
Db	1484	CGCTAAAGTTGGGTGAGTATGGTGAGGTTACGGTTGATTGTGAGCCACGGTCAGGAATAG	1543
Qy	601	ACACTAGCGCTTACTACGTTATGTGAGTGGGTGCGAAGTCCTTCTTGGTTACCCGAGAAT	660
Db	1544	ACACCAGCGCCTATTACGTTATGTGAGTGGTGAGAAGTCCTTCTTGGTTACCCGAGAAT	1603
Qy	661	GGTTTATGGACCTGAACCTTCCATGGAGTAGCGCTGGAAGCACAACTGGAGGAACCGGG	720
Db	1604	GGTTTATGGATCTGAACCTGCCATGGAGCAGTGCTGGAAGCACCACTGGAGGAACCGGG	1663
Qy	721	AAACACTGATGGAGTTTGAAGAACCTCATGCCACCAAACAATCTGTCTGAGTCTAGGGT	780
Db	1664	AAACACTGATGGAGTTTGAAGAACCTCATGCCACCAAACAATCTGTTGTGGCTCTAGGGT	1723
Qy	781	CGCAGGAAGGTGCCTTGACCAAGCTCTGGCTGGAGCAATTCCTGTTGAGTTCTCAAGCA	840
Db	1724	CGCAGGAAGGTGCGTTGCACCAAGCTCTGGCCGGAGCGATTTCCTGTTGAGTTCTCAAGCA	1783
Qy	841	ACACTGTGAAGTTGACATCAGGACATCTGAAGTGTAGGGTGAAGATGGAGAAGTTGCAGC	900
Db	1784	ACACTGTGAAGTTGACATCAGGACATCTGAAGTGTGCGGTGAAGATGGAGAAGTTGCAGC	1843

Qy 901 TGAAGGGAACAACATATGGTGTATGCTCAAAAGCATTCAAATTCGCTAGGACTCCCCGCTG 960
 |||
 Db 1844 TGAAGGGAACAACATATGGAGTATGTTCAAAAGCGTTCAAATTCGCTAGGACTCCCCGCTG 1903
 Qy 961 ACACTGGTCATGGAACGGTGGTGTGCTGGAAGTGCAGTATACCGGAAAAGACGGGCCTTGCA 1020
 |||
 Db 1904 ACACTGGCCACGGAACGGTGGTGTGGAAGTGAATATACCGGAACAGACGGTCCCTGCA 1963
 Qy 1021 AAGTGCCCATTTCTTCTGTGGCTTCCCTGAACGACCTTACACCCGTTGGAAGGCTGGTGA 1080
 |||
 Db 1964 AAGTGCCCATTTCTTCCGTAGCTTCCCTGAATGACCTCACACCTGTTGGAAGACTGGTGA 2023
 Qy 1081 CTGTGAATCCATTTGTGTCTGTGGCTACGGCCAACTCGAAGGTTTTGATTGAACTCGAAC 1140
 | |||
 Db 2024 CCGTGAATCCATTTGTGTCTGTGGCCACAGCCAACTCGAAGGTTTTGATTGAACTCGAAC 2083
 Qy 1141 CCCCCTTTAGTGACTCTTACATCGTGGTGGGAGAGGAGAACAGCAGATAAACCACCACT 1200
 |||
 Db 2084 CCCCCTTTGGTGACTCTTACATCGTGGTGGGAAGAGGAGAACAGCAGATAAACCATCACT 2143
 Qy 1201 GGCACAAATCTGGGAGCAGTATTGGAAAGGCTTTACCACTACACTCAGAGGAGCTCAAC 1260
 |||
 Db 2144 GGCACAAATCTGGGAGCAGCATTGGAAAGGCTTTACCACCACACTCAGAGGAGCTCAAC 2203
 Qy 1261 GACTTGCAGCTCTTGAGACACTGCCTGGGATTTTGGATCAGTCGGAGGGGTTTTCACCT 1320
 |||
 Db 2204 GACTCGCAGCTCTTGAGATACTGCTTGGGATTTTGGATCAGTTGGAGGGGTTTTCACCT 2263
 Qy 1321 CGGTAGGGAAAGCCATACACCAAGTTTTTGGAGGAGCCTTTAGATCACTCTTTGGAGGGA 1380
 | |||
 Db 2264 CAGTGGGGAAAGCCATACACCAAGTCTTTGGAGGAGCCTTTAGATCACTCTTTGGAGGGA 2323
 Qy 1381 TGTCTTGGATCACACAGGGGCTTCTGGGAGCTCTTCTGCTGTGGATGGGAATTAACGCCC 1440
 |||
 Db 2324 TGTCTTGGATCACACAGGACTTCTGGGAGCTCTTCTGTTGTGGATGGGAATCAATGCCC 2383
 Qy 1441 GTGACAGGTCAATTGCTATGACGTTCTTGCAGTTGGAGGAGTCTTGTCTTCTCTTTCGG 1500
 |||
 Db 2384 GTGACAGGTCAATTGCTATGACGTTCTTGCAGTTGGAGGAGTTTTGCTCTTCTTTCGG 2443
 Qy 1501 TCAACGTCCATGCTG 1515
 |||
 Db 2444 TCAACGTCCATGCTG 2458

RESULT 7

AR365300

LOCUS AR365300 1491 bp DNA linear PAT 03-SEP-2003

DEFINITION Sequence 3 from patent US 5486473.

ACCESSION AR365300

VERSION AR365300.1 GI:34428831

KEYWORDS .

SOURCE Unknown.

ORGANISM Unknown.

Unclassified.

REFERENCE 1 (bases 1 to 1491)

AUTHORS Fujita,H., Yoshida,I., Takagi,M., Manabe,S. and Fukai,K.

TITLE A DNA coding for a Flavivirus antigen
JOURNAL Patent: US 5486473-A 3 23-JAN-1996;
The Research Foundation for Microbial Diseases of Osaka University;
Osaka;
JPX;

FEATURES Location/Qualifiers
source 1. .1491
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Query Match 86.3%; Score 1312.2; DB 2; Length 1491;
Best Local Similarity 92.7%; Pred. No. 0;
Matches 1393; Conservative 0; Mismatches 98; Indels 12; Gaps 1;

```
Qy      12 TTCAACTGTTTAGGAATGAGCAACAGGGACTTCCTGGAGGGAGTGTCTGGAGCTACATGG 71
          |||||
Db      1 TTCAACTGTTTAGGAATGAGTAACAGAGACTTCCTGGAGGGAGTGTCTGGAGCTACATGG 60

Qy      72 GTTGATCTGGTACTGGAAGGAGACAGTTGTGTGACCATAATGTCAAAAGACAAGCCAACC 131
          |||||
Db      61 GTTGATCTGGTACTGGAAGGCGATAGTTGTGTGACCATAATGTCAAAAGACAAGCCAACC 120

Qy     132 ATTGATGTCAAAATGATGAACATGGAAGCAGCTAATCTCGCAGATGTGCGTAGCTACTGC 191
          |||||
Db     121 ATTGATGTCAAAATGATGAACATGGAAGCAGCCAACCTCGCAGATGTGCGCAGTTACTGT 180

Qy     192 TACTTAGCTTCGGTCAGTGATCTGTCAACAAAAGCCGCGTGTCCAACCATGGGTGAAGCT 251
          |||
Db     181 TACCTAGCTTCGGTCAGTGACTTGTCAACAAGAGCTGCGTGTCCAACCATGGGTGAAGCC 240

Qy     252 CACAACGAGAAAAGAGCCGACCCTGCCTTTGTTTGCAAGCAAGGCGTCGTAGACAGAGGA 311
          |||||
Db     241 CACAACGAGAAAAGAGCTGACCCCGCCTTCGTTTGCAAGCAAGGCGTTGTGGACAGAGGA 300

Qy     312 TGGGGGAATGGATGCGGACTGTTTGGAAAGGGGAGCATTGACACATGTGCAAAGTTTGCC 371
          |||||
Db     301 TGGGGGAATGGCTGCGGACTGTTTGGAAAGGGGAGCATTGACACATGTGCGAAGTTTGCC 360

Qy     372 TGTACAACCAAGGCAACTGGTTGGATTATCCAGAAGGAAAACATCAAGTACGAGGTTGCC 431
          |||||
Db     361 TGTACAACCAAGCAACTGGATGGATCATCCAGAAGGAAAACATCAAGTATGAGGTTGCC 420

Qy     432 ATATTTGTGCATGGCCCGACGACTGTGCAATCACATGGCAATTATTCAACACAGATAGGG 491
          |||||
Db     421 ATATTTGTGCATGGCCCGACGACCGTTGAATCTCATGGCA-----AGATAGGG 468

Qy     492 GCTACCCAAGCAGGAAGGTTTCAAGCATAACTCCATCGGCACCATCCTACACGCTGAAGTTG 551
          ||
Db     469 GCCACCCAGGCTGGAAGATTCAAGTATAACTCCATCGGCACCATCTTACACGCTAAAGTTG 528

Qy     552 GGTGAGTATGGTGAGGTCACAGTTGACTGTGAGCCACGGTCAGGAATAGACACTAGCGCT 611
          |||||
Db     529 GGTGAGTATGGTGAGGTTACGGTTGATTGTGAGCCACGGTCAGGAATAGACACCAGCGCC 588

Qy     612 TACTACGTTATGTCAAGTGGGTGCGAAGTCCTTCTTGGTTACCGAGAATGGTTTATGGAC 671
          ||
```

Db	589	TATTACGTTATGTCTAGTTGGTGAGAACTCCTTCCTGGTTACCGGAGAATGGTTTATGGAT	648
Qy	672	CTGAACCTTCCATGGAGTAGCGCTGGAAGCACAACTGGAGGAACCGGGAAACACTGATG	731
Db	649	CTGAACCTGCCATGGAGCAGTGCTGGAAGCACCACGTGGAGGAACCGGGAAACACTGATG	708
Qy	732	GAGTTTGAAGAACCTCATGCCACCAAACAATCTGTCTAGCTCTAGGGTCGCAGGAAGGT	791
Db	709	GAGTTTGAAGAACCTCATGCCACCAAACAATCTGTTGTGGCTCTAGGGTCGCAGGAAGGT	768
Qy	792	GCCTTGCACCAAGCTCTGGCTGGAGCAATTCTGTTGAGTTCTCAAGCAACACTGTGAAG	851
Db	769	GCGTTGCACCAAGCTCTGGCCGGAGCGATTCTGTTGAGTTCTCAAGCAACACTGTGAAG	828
Qy	852	TTGACATCAGGACATCTGAAGTGTAGGGTGAAGATGGAGAAGTTGCAGCTGAAGGGAACA	911
Db	829	TTGACATCAGGACATCTGAAGTGTCTGGGTGAAGATGGAGAAGTTGCAGCTGAAGGGAACA	888
Qy	912	ACATATGGTGTATGCTCAAAAGCATTCAAATTCGCTAGGACTCCCGCTGACACTGGTCAT	971
Db	889	ACATATGGAGTATGTTCAAAAGCGTTCAAATTCGCTAGGACTCCCGCTGACACTGGCCAC	948
Qy	972	GGAACGGTGGTGCTGGAAGTGCAGTATACCGGAAAAGACGGGCCTTGCAAAGTGCCCAT	1031
Db	949	GGAACGGTGGTGTTGGAAGTGAATATACCGGAACAGACGGTCCCTGCAAAGTGCCCAT	1008
Qy	1032	TCTTCTGTGGCTTCCCTGAACGACCTTACACCCGTTGGAAGGCTGGTGACTGTGAATCCA	1091
Db	1009	TCTTCCGTAGCTTCCCTGAATGACCTCACACCTGTTGGAAGACTGGTGACCGTGAATCCA	1068
Qy	1092	TTTGTGTCTGTGGCTACGGCCAACTCGAAGGTTTTGATTGAACTCGAACCCCGTTTAGT	1151
Db	1069	TTTGTGTCTGTGGCCACAGCCAACTCGAAGGTTTTGATTGAACTCGAACCCCGTTTGGT	1128
Qy	1152	GACTCTTACATCGTGGTGGGGAGAGGAGAACAGCAGATAAACCACCACTGGGCACAAATCT	1211
Db	1129	GACTCTTACATCGTGGTGGGAAGAGGAGAACAGCAGATAAACCATCACTGGGCACAAATCT	1188
Qy	1212	GGGAGCAGTATTGGAAGGCTTTCACCACTACACTCAGAGGAGCTCAACGACTTGCAGCT	1271
Db	1189	GGGAGCAGCATTGGAAGGCCTTTACCACCACACTCAGAGGAGCTCAACGACTCGCAGCT	1248
Qy	1272	CTTGGAGACACTGCCTGGGATTTTGGATCAGTCGGAGGGGTTTTCACCTCGGTAGGGAAA	1331
Db	1249	CTTGGAGATACTGCTTGGGATTTTGGATCAGTTGGAGGGGTTTTCACCTCAGTGGGGAAA	1308
Qy	1332	GCCATACACCAAGTTTTTGGAGGAGCCTTTAGATCACTCTTTGGAGGGATGTCTGGATC	1391
Db	1309	GCCATACACCAAGTCTTTGGAGGAGCTTTTAGATCACTCTTTGGAGGGATGTCTGGATC	1368
Qy	1392	ACACAGGGGCTTCTGGGAGCTCTTCTGCTGTGGATGGGAATTAACGCCCGTGACAGGTCA	1451
Db	1369	ACACAGGGACTTCTGGGAGCTCTTCTGTTGTGGATGGGAATCAATGCCCGTGACAGGTCA	1428
Qy	1452	ATTGCTATGACGTTCTTGCAGTTGGAGGAGTCTTGCTCTTCTTTTCGGTCAACGTCCAT	1511
Db	1429	ATTGCTATGACGTTTCTTGCAGTTGGAGGAGTTTTGCTCTTCTTTTCGGTCAACGTCCAT	1488

Qy 1512 GCT 1514
 | | |
Db 1489 GCT 1491

RESULT 8

AF394221

LOCUS AF394221 1430 bp mRNA linear VRL 03-MAY-2002

DEFINITION West Nile virus isolate B956 polyprotein mRNA, envelope
glycoprotein E and nonstructural protein 1 region, partial cds.

ACCESSION AF394221

VERSION AF394221.1 GI:20428494

KEYWORDS .

SOURCE West Nile virus

ORGANISM West Nile virus

Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Flavivirus; Japanese encephalitis virus group.

REFERENCE 1 (bases 1 to 1430)

AUTHORS Briese,T., Rambaut,A., Pathmajeyan,M., Bishara,J., Weinberger,M.,
Pitlik,S. and Lipkin,W.I.

TITLE Phylogenetic analysis of a human isolate from the 2000 Israel West
Nile virus epidemic

JOURNAL Emerging Infect. Dis. 8 (5), 528-531 (2002)

PUBMED 11996693

REFERENCE 2 (bases 1 to 1430)

AUTHORS Briese,T., Jordan,I., Pathmajeyan,M. and Lipkin,W.I.

TITLE Direct Submission

JOURNAL Submitted (21-JUN-2001) Emerging Diseases Laboratory, Microbiology
& Molecular Genetics, and Neurology, University California Irvine,
3107 Gillespie Neuroscience Building, Irvine, CA 92697-4292, USA

FEATURES Location/Qualifiers

source

1. .1430

/organism="West Nile virus"

/mol_type="mRNA"

/isolate="B956"

/db_xref="taxon:11082"

/country="Uganda"

/note="isolated from human serum in 1937; kindly provided
by Bob Tesh, University of Texas Medical Br., Galveston,
TX, USA"

CDS

<1. .>1430

/codon_start=3

/product="polyprotein"

/protein_id="AAM21944.1"

/db_xref="GI:20428495"

/translation="KRADPAFVCKQGVVDRGWGNGCGLFGKGSIDTCAKFACTTKATG
WIIQKENIKYEVAIFVHGPTTVESHGKIGATQAGRFSITPSAPSYTLKLGEYGEVTV
CEPRSGIDTSAYYVMSVGAKSFLVHREWFMDLNLWPSSAGSTTWRNRETIVEFEPEHA
TKQSVVALGSQEGALHQALAGAIPEFSSNTVKLTSGHLKCRVKMEKLQLKGTITYGVC
SKAFKFARTPADTGHGTVVLELQYTGTGDPCKVPISSVASLNDLTPVGRLVTVNPFVS
VATANSKVLIELEPPFGDSYIVVGRGEQQINHHWHKSGSSIGKAFTTTLRGAQRLAAL
GDTAWDFGSVGGVFTSVGKAHQVFGGAFRSLFGGMSWITQGLLGALLLWGMGINARDR
SIAMTFLAVGGVLLFLSVNVHADTGCAIDIGRQELRCGSGVFIHNDVEAWMDRYKFYP
ETPQGLAKIIQKAHAEGVCGLRSVSR"

mat_peptide

<1. .1244

/product="envelope glycoprotein E"

ORIGIN

Qy		259 AGAAAAGAGCCGACCCTGCCTTTGTTCGAAGCAAGGCCGTCGTACAGAGGATGGGGGA	318
Db	1	AGAAAAGAGCTGACCCCGCCTTCGTTTGCAAGCAAGGCCGTTGTGGACAGAGGATGGGGAA	60
Qy		319 ATGGATGCGGACTGTTTTGGAAAGGGGAGCATTTGACACATGTGCAAAGTTTGCCTGTACAA	378
Db	61	ATGGCTGCGGACTGTTTTGGAAAGGGGAGCATTTGACACATGTGCGAAGTTTGCCTGTACAA	120
Qy		379 CCAAGGCAACTGGTTGGATTATCCAGAAGGAAAAACATCAAGTACGAGGTTGCCATATTTG	438
Db	121	CCAAAGCAACTGGATGGATCATCCAGAAGGAAAAACATCAAGTATGAGGTTGCCATATTTG	180
Qy		439 TGCATGGCCCCGACGACTGTCGAATCACATGGCAATTATTCAACACAGATAGGGGCTACCC	498
Db	181	TGCATGGCCCCGACGACCGTTGAATCTCATGGCA-----AGATAGGGGCCACCC	228
Qy		499 AAGCAGGAAGGTTTCAGCATAACTCCATCGGCACCATCCTACACGCTGAAGTTGGGTGAGT	558
Db	229	AGGCTGGAAGATTCA GTATAACTCCATCGGC GCCATCTTACACGCTAAAGTTGGGTGAGT	288
Qy		559 ATGGTGAGGTCACAGTTGACTGTGAGCCACGGTCAGGAATAGACACTAGCGCTTACTACG	618
Db	289	ATGGTGAGGTTACGTTGATTGTGAGCCACGGTCAGGAATAGACACTAGCGCCTATTACG	348
Qy		619 TTATGTCAGTGGGTGCGAAGTCCTTCTTGGTTCACCGAGAATGGTTTATGGACCTGAACC	678
Db	349	TTATGTCAGTTGGTGCGAAGTCCTTCTTGGTTCACCGAGAATGGTTTATGGATCTGAACC	408
Qy		679 TTCCATGGAGTAGCGCTGGAAGCAC AACGTGGAGGAACCGGAAACACTGATGGAGTTTG	738
Db	409	TGCCATGGAGCAGTGCTGGAAGCACACGTGGAGGAACCGGAAACACTGGTGGAGTTTG	468
Qy		739 AAGAACCTCATGCCACCAAACAATCTGTCGTAGCTCTAGGGTCGCAGGAAGGTGCCTTGC	798
Db	469	AAGAACCTCATGCCACCAAACAATCTGTTGTGGCTCTAGGGTCGCAGGAAGGTGCCTTGC	528
Qy		799 ACCAAGCTCTGGCTGGAGCAATTCCTGTTGAGTTCTCAAGCAACACTGTGAAGTTGACAT	858
Db	529	ACCAAGCTCTGGCCGGAGCGATTTCCTGTTGAGTTCTCAAGCAACACTGTGAAGTTGACAT	588
Qy		859 CAGGACATCTGAAGTGTAGGGTGAAGATGGAGAAGTTGCAGCTGAAGGGAACAACATATG	918
Db	589	CAGGACATCTGAAGTGTAGGGTGAAGATGGAGAAGTTGCAGCTGAAGGGAACAACATATG	648
Qy		919 GTGTATGCTCAAAAGCATTCAAATT CGCTAGGACTCCC GCTGACACTGGTCATGGAACGG	978
Db	649	GAGTATGTTCAAAGCGTTCAAATT CGCTAGGACTCCC GCTGACACTGGCCACGGAACGG	708

Qy	979	TGGTGTCTGGAAGCTGCAGTATACCGGAAAAGACGGGCCCTTGCAAAGTGCCCATTTCTTCTG	1038
Db	709	TGGTGTCTGGAAGCTGCAGTATACCGGAAAAGACGGGCCCTTGCAAAGTGCCCATTTCTTCTG	768
Qy	1039	TGGCTTCCCTGAACGACCTTACACCCGTTGGAAGGCTGGTGACTGTGAATCCATTTGTGT	1098
Db	769	TAGCTTCCCTGAATGACCTCACACCTGTTGGAAGACTGGTGACCGTGAATCCATTTGTGT	828
Qy	1099	CTGTGGCTACGGCCAACTCGAAGGTTTTGATTGAACTCGAACCCCCGTTTAGTGACTCTT	1158
Db	829	CTGTGGCCACAGCCAACTCGAAGGTTTTGATTGAACTCGAACCCCCGTTTGGTGACTCTT	888
Qy	1159	ACATCGTGGTGGGAGAGGAGAACAGCAGATAAACCACTGGCACAAATCTGGGAGCA	1218
Db	889	ACATCGTGGTGGGAAGAGGAGAACAGCAGATAAACCATCACTGGCACAAATCTGGGAGCA	948
Qy	1219	GTATTGGAAAGGCTTTTACCACCTACACTCAGAGGAGCTCAACGACTTGCAGCTCTTGGAG	1278
Db	949	GCATTGGAAAGGCTTTTACCACCACACTCAGAGGAGCTCAACGACTCGCAGCTCTTGGAG	1008
Qy	1279	ACACTGCCTGGGATTTTGGATCAGTCGGAGGGGTTTTACCTCGGTAGGGAAAGCCATAC	1338
Db	1009	ATACTGCTTGGGATTTTGGATCAGTTGGAGGGGTTTTACCTCAGTGGGGAAAGCCATAC	1068
Qy	1339	ACCAAGTTTTTGGAGGAGCCTTTAGATCACTCTTGGAGGGATGTCTCGGATCACACAGG	1398
Db	1069	ACCAAGTCTTTGGAGGAGCCTTTAGATCACTCTTGGAGGGATGTCTCGGATCACACAGG	1128
Qy	1399	GGCTTCTGGGAGCTCTTCTGCTGTGGATGGGAATTAACGCCCGTGACAGGTCAATTGCTA	1458
Db	1129	GACTTCTGGGAGCTCTTCTGTTGTGGATGGGAATCAATGCCCGTGACAGGTCAATTGCTA	1188
Qy	1459	TGACGTTTCCTTGCGGTTGGAGGAGTCTTGCTCTTCCTTTCGGTCAACGTCCATGCTG	1515
Db	1189	TGACGTTTCTTGCGGTTGGAGGAGTTTTGCTCTTCCTTTCGGTCAACGTCCATGCTG	1245

TITLE Direct Submission

JOURNAL Submitted (29-JUL-2004) CNR Arbovirus, Institut Pasteur, 21 Avenue
Tony Garnier, Lyon cedex 07 69365, France

FEATURES Location/Qualifiers
source 1. .10945
/organism="West Nile virus"
/mol_type="genomic RNA"
/strain="04.05"
/isolation_source="brain of horse with encephalitis"
/specific_host="horse"
/db_xref="taxon:11082"
/country="Morocco"
/collection_date="2003"
CDS 55. .10356
/codon_start=1
/product="polyprotein"
/protein_id="AAT92099.1"
/db_xref="GI:51011376"
/translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
KKRGGKTGIAVMIGLIASVGAVTLSNFQGGKVMVTNATDVTDVITPTAAGKNLCIVR
AMDVGVMCDDTITYECPVLSAGNDPEDIDCWCTKSAYVVRGCTKTRHSRRSRRSLT
VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
VFVVLVLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRAPAFVCRQGVVDRGWG
NGCGLFGKGSIDTCAKFACSTKATGRTILKENIKYEVAIFVHGPTTVESHGNYSTQIG
ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
MDLNLWPSSAGSTVWRNRETLMFEFEHPHATKQSVIALGSGEGALHQALAGAI PVEFSS
NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
PCKVPISSVASLNDLTPVGRLVTNPFVSVATANAKVLIIELEPPFGDSYIVVGRGEQQ
INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSGGVFTSVGKAVHQVFGGAF
RSLFGGMSWITQGLLGAALLWGMINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
ISRQELRCGSGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
QMWESVKDELNTLLKENGVDLSVVVEKQEGMYKSAPKRLTATTEKLEIGKAWGKSIL
FAPELANNTFVVDGPETKECPTQNRANSLEVEDFGFGFLTSTRMFLKVRESNTTECDS
KIIIGTAIKNNLAIHSDLSYWIESRLNDTWKLERAVLGEVKSTWPTHTLWGDGILES
DLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSSECGHRGPA
TRTTTESGKLITDWCCRSCTLPLRLYQTDSCGWYGMIEIRPQRHDEKTLVQSQVNAYNA
DMIDPFQLGLLVFLATQEVLRKRWTAKISMPAILIALLVLFVGGITYTDVLRVILV
GAFAESNSGGDVVHLALMATFKIQPFVFMVASFLKARWTNQENILLMLAAVFFQMAYY
DARQILLWEIPDVNLAVAWMILRAITFTTTSNVVPLALLTPGLRCLNLDVYRIL
LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMLAAGLIACDPNRKRGWPA
TEVMTAVGLMFAIVGGLAELDIDSMAPMTIAGLMFAAFVISGKSTDMWIERTADISW
ESDAEITGSSSERVDVRLDDDNFQLMNDPGAPWKIWMMLRMACLAISAYTPWAILPSVV
GFWITLQYTKRGGVLWDTPSPKEYKKGDTTGVYRIMTRGLLGSYQAGAGVMVEGVLH
TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMI VVEPGKN
VKNVQTKPGVFKTPEGEIGAVTLDFTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
AIVQGERMDEPI PAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
LAPTRVVAEAEALRGLPIRYQTSAVTREHNGNEIVDMCHATLTHRLMSPHRVPNY
NLFVMDAEHFTDPASIAARGYISTKVELGEAAAFMTATPPGTSDFPESNSPISDLQ
TEIPDRAWNSGYEWITEYIGKTVWFVPSVKMGNEIALCLQRAGKKVVQLNRKSYET
PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEP
SAVTAASAAQRRGRIGRNPQVQVGEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLIAQFYQ
PEREKVYTM DGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVSYHRRWCFDGP
R TNILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALLSVMTMG
VFFLLMQRKIGIGKIGLGGVVLGVATFFCWMAEVPGTKIAGMLLLSLLMIVLIP
PEPEKQRSQTDNQLAVFLICVMTLVSAVAANEMGWLDKTKSDISSLFGQRIE
VKENFSMGEFL

Qy	481	CACAGATAGGGGCTACCCAAGCAGGAAGGTTGAGCATAACTCCATCGGCACCATCCTACA	540
Db	1394	CACAGATTGGGGCCACTCAGGCAGGGAGATTGAGCATCACTCCTGCGGCGCCTTCATACA	1453
Qy	541	CGCTGAAGTTGGGTGAGTATGGTGAGGTCACAGTTGACTGTGAGCCACGGTCAGGAATAG	600
Db	1454	CACTAAAGCTTGGAGAATATGGAGAAGTGACAGTGGACTGTGAACCACGGTCAGGGATTG	1513
Qy	601	ACACTAGCGCTTACTACGTTATGTCAGTGGGTGCGAAGTCCTTCTTGGTTACCCGAGAAT	660
Db	1514	ACACCAATGCCTACTACGTGATGACTGTTGGAACAAAGACGTTTTTGGTCCATCGTGAGT	1573
Qy	661	GGTTTATGGACCTGAACCTTCCATGGAGTAGCGCTGGAAGCACAACGTGGAGGAACCGGG	720
Db	1574	GGTTCATGGACCTCAACCTCCCTTGGAGCAGTGCTGGAAGTACTGTGTGGAGGAATAGAG	1633
Qy	721	AAACACTGATGGAGTTTGAAGAACCTCATGCCACCAACAATCTGTCTAGCTCTAGGGT	780
Db	1634	AGACGTTAATGGAGTTTGAAGAACACACGCCACAAAGCAGTCTGTGATAGCACTGGGCT	1693
Qy	781	CGCAGGAAGGTGCCTTGACCAAGCTCTGGCTGGAGCAATTCCTGTTGAGTTCTCAAGCA	840
Db	1694	CACAAGAGGGAGCTCTGCATCAAGCTTTGGCTGGAGCCATCCCTGTGGAATTTTCAAGCA	1753
Qy	841	ACACTGTGAAGTTGACATCAGGACATCTGAAGTGTAGGGTGAAGATGGAGAAGTTGCAGC	900
Db	1754	ACACTGTTAAGTTGACGTCGGGTCATCTGAAGTGTAGAGTGAAGATGGAAAATTGCAGT	1813
Qy	901	TGAAGGGAACAACATATGGTGTATGCTCAAAGCATTCAAATTGCTAGGACTCCCCTG	960
Db	1814	TGAAGGGAACAACCTACGGCGTCTGTTCAAAGGCTTTCAAGTTTCTTGGGACTCCCGCAG	1873
Qy	961	ACACTGGTCATGGAACGGTGGTGTGGAAGTGCAGTATACCGGAAAAGACGGGCCTTGCA	1020
Db	1874	ACACAGGTCACGGCACTGTGGTGTGGAATTGCAGTACACTGGCACGGATGGACCTTGCA	1933
Qy	1021	AAGTGCCCATTTCTTCTGTGGCTTCCCTGAACGACCTTACACCCGTTGGAAGGCTGGTGA	1080
Db	1934	AAGTTCCCATCTCGTCAGTGGCTTCTTGAACGACCTAACACCGGTGGGCAGATTGGTCA	1993
Qy	1081	CTGTGAATCCATTTGTGTCTGTGGCTACGGCCAACCTCGAAGGTTTTGATTGAACTCGAAC	1140
Db	1994	CTGTTAACCCTTTTGTTCAGTGGCCACGGCCAATGCCAAGGTCCTGATTGAACTGGAAC	2053
Qy	1141	CCCCGTTTGTAGTACTCTTACATCGTGGTGGGGAGAGGAGAACAGCAGATAAACCACCACT	1200
Db	2054	CACCTTTTGGAGACTCATACATAGTGGTAGGCAGAGGAGAACAACAGATCAATCACCATT	2113
Qy	1201	GGCACAAATCTGGGAGCAGTATTGGAAAGGCTTTCAACCACTACACTCAGAGGAGCTCAAC	1260
Db	2114	GGCATAAGTCTGGAAGCAGCATCGGCAAAGCCTTTACAACCACTCTCAAAGGGGCGCAGA	2173
Qy	1261	GACTTGCAGCTCTTGGAGACACTGCCTGGGATTTTGGATCAGTCGGAGGGGTTTTACCT	1320
Db	2174	GATTAGCCGCTCTAGGAGACACAGCTTGGGACTTTGGATCAGTTGGAGGGGTGTTACCT	2233

Qy	1321	CGGTAGGGAAAGCCATACACCAAGTTTTTGGAGGAGCCTTTAGATCACTCTTTGGAGGGA	1380
Db	2234	CAGTAGGGAAGGCTGTCCATCAAGTGTCGGTGGAGCGTTCGCTCACTGTTTGGAGGTA	2293
Qy	1381	TGTCCTGGATCACACAGGGGCTTCTGGGAGCTCTTCTGCTGTGGATGGGAATTAACGCCC	1440
Db	2294	TGTCCTGGATAACGCAGGGATTGCTGGGGGCTCTTCTGTTGTGGATGGGCATCAATGCTC	2353
Qy	1441	GTGACAGGTCAATTGCTATGACGTTCTTGCGGTTGGAGGAGTCTTGCTCTTCCTTTCGG	1500
Db	2354	GTGACAGGTCCATAGCTCTCACGTTTCTCGCAGTTGGAGGAGTCTGCTCTTCCTCTCCG	2413
Qy	1501	TCAACGTCCATGCTG	1515
Db	2414	TGAACGTGCACGCTG	2428

RESULT 10

AY262283

LOCUS	AY262283	10984 bp	RNA	linear	VRL 29-OCT-2003
-------	----------	----------	-----	--------	-----------------

DEFINITION West Nile virus isolate KN3829 polyprotein gene, complete cds.

ACCESSION AY262283

VERSION AY262283.1 GI:30230630

KEYWORDS

SOURCE West Nile virus (WNV)

ORGANISM West Nile virus

Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Flavivirus; Japanese encephalitis virus group.

REFERENCE 1 (bases 1 to 10984)

AUTHORS Charrel, R.N., Brault, A.C., Gallian, P., Lemasson, J.-J., Murgue, B., Murri, S., Pastorino, B., Zeller, H., de Chesse, R., de Micco, P. and de Lamballerie, X.

TITLE Evolutionary relationship between Old World West Nile virus strains. Evidence for viral gene flow between africa, the middle east, and europe

JOURNAL Virology 315 (2), 381-388 (2003)

PUBMED 14585341

REFERENCE 2 (bases 1 to 10984)

AUTHORS Brault, A.C. and de Lamballerie, X.

TITLE Direct Submission

JOURNAL Submitted (25-MAR-2003) Division of Vector-Borne Infectious
Diseases, Centers for Disease Control and Prevention, P.O. Box
2087, Fort Collins, CO 80522, USA

FEATURES	Location/Qualifiers
----------	---------------------

source

1. .10984

```
/organism="West Nile virus"
```

```
/mol type="genomic RNA"
```

```
/isolate="KN3829"
```

```
/specific host="Culex univittatus"
```

```
/db xref="taxon:11082"
```

5'UTR 1.60

CDS	61.10362
-----	----------

```
/codon start=1
```

```
/product="polyprotein"
```

```
/protein id="AAP20887.1"
```

```
/db xref="GI:30230631"
```

/translation="MSKKPGGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI

RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLISFKKELGTITSAINRRSSKQ
KKRGNTGIAAMIGLIASVGAVTILSNFQQKVMMTVNATDVTDVITPTAAGKNLCIVR
AMDVGVMCDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPYGALVAAVIGWMLGSNTMQRV
VFVVLVLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
MMNMEANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWG
NGCGLFGKGSIDTCAKFACSTKATGRTILKENIKYEVAIFVHGPTTVESHGNYSTQIG
ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
MDLNLFWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPEVFSS
NTVKLTSGHLKCRVKMEKLQKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
PCKVPISSVASLNDLTPVGRVTVNPFVSVATANAKVLIIELEPPFGDSYIVVGRGEQQ
INHHWHKSGSSIGKAFTTTTLKGAQRLAALGDTAWDFGSGVGVFTSVGKAVHQVFGGAF
RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
ISRQELRCGSGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
QMWESVKDELNTLLKENGVDLSVVVEKQEGMYKSAPKRLTATTEKLEIGKAWGKSIL
FAPELANNTFVVDGPETKECPTQNRWNSLEVEDFGFGLTSTRMFLKVRESNTTECD
KII GTAVKNNLAIHSDLSYWIESRLNDTWKLERAVLGEVKSCTWPETHTLWGDGILES
DLIIPVTLAGPRSNHNRRPGYKTQNGQPWDEGRVEIDFDYCPGTTVTLSSECGHRGPA
TRTTTESGKLITDWCCRSC TLPPLRYQTDSCGWYMEIRPQRHDEKTLVQSQVNAAYNA
DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLFVGGITYADVLRYVILV
GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILMLAAVFFQ MAYH
DARQILLWEIPDVLSLAVAWMILRAITFTTTSNVVVPLLALLTPGLRCLNLDVYRIL
LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRGWPA
TEVMTAVGLMFAIVGGLAELDIDSMAPMTIAGLMFAAFVISGKSTDMWIERTADISW
ESDAEITGSSSERVDVRLDDDGNFQLMNDPGAPWKIWMRLMACLAISAYTPWAILPSV
GFWITLQYTKRGGVLWDTPSPKEYKKGDTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQMI VVEPGKN
VKNVQTKPGVFKTPEGEIGAVTLDFTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
AIVQGERMDEPI PAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRRLTAV
LAPTRVVAEAMAEALRGLPIRYQTSAVTREHNGNEIVDMCHATLTHRLMSPHRVPNY
NLFVMDAEAHFTDPASIAARGYISTKVELGEAAAFMTATPPGTSDFPFESNSPISDLQ
TEIPDRAWNSGYEWITEYIGKT VWFVPSVKMGNEIALCLQ RAGKKVQ LNRKSYET EY
PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAA
SAAQRRGRI GRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLIAQFYQ
PEREKVYTM DGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVSYHRRWC F D GPR
TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALLIALLSVMTMG
VFFLLMQRKIGIKIGLGGVVLGVATFFCWMAEVPGTKIAGMLLLSLLMIVLIP EPEK
QRSQTDNQLAVFLICVMTLVSAVANEMGWLDKTKSDISSLFGQRIEVKENFSMGEFL
LDLRPATAWSLYAVTTAVLTPLLKHLITS DYINTSLTSINVQASALFTLARGFPFVDV
GVSALLLAAGCWGQVTLTVTVAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
VVDGIVATDVPELERTTPIMQKKVQIMLILVSLAAVVVNPSVKTVREAGILITAAAV
TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLIKNMDKPKLKRGGAKGRTLGE
VVKERLNQMTKEEFTRYRKEAII EVDRSAAKHARKEGNVTGGHPVSRGTAKLRWLVER
RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFCVKVL
CPYMPKVI EK MELLQRRYGGGLVRNPLSRNSTHEMYWVSASGNVVSVMNTSQVLLG
RMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLRREYSSTWHHDENH
PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDITITNVTMTAMTDTTPFGQQRVFK
EKVDTKAPEPPEGVKYVLNETTNWLWAF LAREKRPRMCSREEFIRKVN SNAALGAMFE
EQNQWRSAREAVEDPKFWEMVDEERE AHLRGECHTCIYNMMGKREKKPGFEGKAKGSR
AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLG YILREVGT RP GGI
YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAI IELTYRHKVVKVMRPAADGR
TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTGKGKPKV
RTWLFENGEE RLSRMAVSGDDCVVKPLDDRFATSLHFLNAMS KVRKDIQEWKPSTGWY
DWQQVPFC SNHFTELIMKDGR TLVVP CRGQDELVGRARISPGAGWNVRDTACLA KSYA
QMWLLLYFHRRDLRLMANAICSAVPVNVVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW

```

IEENEWMEDKTPVEKWSDPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRAIIG
DEKYVDYMSSLKRYEDTTLVEDTVL"
mat_peptide      61. .429
                  /product="capsid"
mat_peptide      430. .705
                  /product="prM"
mat_peptide      706. .930
                  /product="M"
mat_peptide      931. .2433
                  /product="envelope"
mat_peptide      2434. .3492
                  /product="NS1"
mat_peptide      3493. .4182
                  /product="NS2A"
mat_peptide      4183. .4575
                  /product="NS2B"
mat_peptide      4576. .6432
                  /product="NS3"
mat_peptide      6433. .6879
                  /product="NS4A"
mat_peptide      6880. .7644
                  /product="NS4B"
mat_peptide      7645. .10359
                  /product="NS5"
3'UTR            10363. .10984
ORIGIN

```

```

Query Match          65.8%; Score 999.8; DB 10; Length 10984;
Best Local Similarity 78.7%; Pred. No. 0;
Matches 1193; Conservative 0; Mismatches 322; Indels 0; Gaps 0;

```

```

Qy      1 CGGAATTCAGCTTCAACTGTTTAGGAATGAGCAACAGGGACTTCCTGGAGGGAGTGTCTG 60
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      920 CAGCCTACAGCTTCAACTGTCTTGAATGAGCAACAGAGACTTCTTGGAGGGAGTATCTG 979

Qy      61 GAGCTACATGGGTTGATCTGGTACTGGAAGGAGACAGTTGTGTGACCATAATGTCAAAAG 120
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      980 GAGCAACATGGGTGGATTTGGTTCTCGAAGGCGACAGCTGCGTGACTATCATGTCTAAGG 1039

Qy      121 ACAAGCCAACCATTGATGTCAAAATGATGAACATGGAAGCAGCTAATCTCGCAGATGTGC 180
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1040 ACAAGCCTACCATTGATGTGAAGATGATGAATATGGAGGCTGCCAACCTGGCAGAGGTCC 1099

Qy      181 GTAGCTACTGCTACTTAGCTTCGGTCAGTGATCTGTCAACAAAAGCCGCGTGTCCAACCA 240
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1100 GCAGTTATTGCTATTTGGCTACCGTCAGCGATCTCTCCACCAAAGCTGCATGCCCGACCA 1159

Qy      241 TGGGTGAAGCTCACAACGAGAAAAGAGCCGACCCTGCCTTTGTTTGCAAGCAAGGCGTCG 300
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1160 TGGGAGAAGCTCACAATGACAAACGTGCTGACCCAGCTTTTGTGTGCAGACAAGGAGTGG 1219

Qy      301 TAGACAGAGGATGGGGGAATGGATGCGGACTGTTTGGAAAGGGGAGCATTGACACATGTG 360
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1220 TGGACAGGGGCTGGGGCAACGGCTGCGGACTATTTGGCAAAGGAAGCATTGACACATGCG 1279

Qy      361 CAAAGTTTGCCTGTACAACCAAGGCAACTGGTTGGATTATCCAGAAGGAAAACATCAAGT 420
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Db	1280	CCAAATTTGCCTGCTCCACCAAGGCCAAGGAAGAACCATCTTGAAAGAGAATATCAAGT	1339
Qy	421	ACGAGGTTGCCATATTTGTGCATGGCCCACGACTGTCGAATCACATGGCAATTATTCAA	480
Db	1340	ATGAAGTGGCCATCTTTGTCCATGGACCAACCACTGTGGAGTCGCATGGAACTACTCCA	1399
Qy	481	CACAGATAGGGGCTACCCAAGCAGGAAGGTTT CAGCATAACTCCATCGGCACCATCCTACA	540
Db	1400	CACAGATTGGGGCCACTCAGGCAGGGAGATT CAGCATCACTCCTGCGGCGCCTTCATACA	1459
Qy	541	CGCTGAAGTTGGGTGAGTATGGTGAGGTCACAGTTGACTGTGAGCCACGGTCAGGAATAG	600
Db	1460	CACTAAAGCTTGGAGAATATGGAGAAGTGACAGTGGACTGTGAACCACGGTCAGGGATTG	1519
Qy	601	ACACTAGCGCTTACTACGTTATGT CAGTGGGTGCGAAGTCCTTCTTGTTTACCAGAGAAT	660
Db	1520	ACACCAATGCTTACTACGTGATGACTGTTGGAACAAAGACGTTTTTGGTCCATCGTGAGT	1579
Qy	661	GGTTTATGGACCTGAACCTTCCATGGAGTAGCGCTGGAAGCACAACGTGGAGGAACCGGG	720
Db	1580	GGTTCATGGACCTCAACCTCCCTTGGAGCAGTGCTGGAAGTACTGTGTGGAGGAACAGAG	1639
Qy	721	AAACACTGATGGAGTTTGAAGAACCTCATGCCACCAAACAATCTGTCTAGCTCTAGGGT	780
Db	1640	AGACGTTAATGGAGTTT GAGGAACCACACGCCACAAAGCAGTCTGTGATAGCATTGGGCT	1699
Qy	781	CGCAGGAAGGTGCCTTGACCAAGCTCTGGCTGGAGCAATTCTGTGAGTTCTCAAGCA	840
Db	1700	CACAAGAGGGAGCTCTGCATCAAGCTTTGGCTGGAGCCATCCCTGTGGAATTTTCAAGCA	1759
Qy	841	ACACTGTGAAGTTGACATCAGGACATCTGAAGTG TAGGGTGAAGATGGAGAAGTTGCAGC	900
Db	1760	ACACTGTCAAGTTGACGTCGGGTCA TTTGAAGTG TAGAGTGAAGATGGAAAAATTGCAGT	1819
Qy	901	TGAAGGGAACAACATATGGTGTATGCTCAAAA GCATTCAAATTCGCTAGGACTCCCCGTG	960
Db	1820	TGAAGGGAACAACCTACGGCGTCTGTTCAAAGGCTTTCAAGTTTCTTGGGACTCCCGCAG	1879
Qy	961	ACACTGGTCATGGAACGGTGGTGCTGGA ACTGCAGTATACCGGAAAAGACGGGCCTTGCA	1020
Db	1880	ACACAGGTCACGGCACTGTGGTGTTGGAATTGCAGTACACTGGTACGGATGGACCTTGCA	1939
Qy	1021	AAGTGCCCATTTCTTCTGTGGCTTCCCTGAACGACCTTACACCCGTTGGAAGGCTGGTGA	1080
Db	1940	AAGTTCCCATCTCGTCAGTGGCTTCATTGAACGACCTAACACCAGTGGGCAGATTGGTCA	1999
Qy	1081	CTGTGAATCCATTTGTGTCTGTGGCTACGGCCA ACTCGAAGGTTTTGATTGAACTCGAAC	1140
Db	2000	CTGTCAACCCTTTTTGTTCAGTGGCCACGGCCA ATGCCAAGGTCCTGATTGAACTGGAAC	2059
Qy	1141	CCCCGTTTAGTGACTCTTACATCGTGGTGGGGAGAGGAGA ACAGCAGATAAACACCACCT	1200
Db	2060	CACCCTTTGGAGACTCATACATAGTGGTAGGCAGAGGAGAACAACAGATCAATCACCATT	2119
Qy	1201	GGCACAAATCTGGGAGCAGTATTGAAAAGGCTTTCA CCACTACACTCAGAGGAGCTCAAC	1260
Db	2120	GGCATAAGTCTGGAAGCAGCATTGGCAAAGCCTTTACA ACCACTCTCAAAGGGGCGCAGA	2179

Qy 1261 GACTTGCAGCTCTTGGAGACACTGCCTGGGATTTTGGATCAGTCGGAGGGGTTTTTCACCT 1320
 || | || ||||| ||||| || ||||| ||||| ||||| ||||| |||||
 Db 2180 GATTAGCCGCTCTAGGAGACACAGCTTGGGACTTTGGATCAGTTGGAGGGGTGTTTCACCT 2239

Qy 1321 CGGTAGGGAAAGCCATACACCAAGTTTTTGGAGGAGCCTTTAGATCACTCTTTGGAGGGA 1380
 | ||||| || | || ||||| || || ||||| || | ||||| || ||||| |
 Db 2240 CAGTAGGGAAGGCTGTCCATCAAGTGTTCGGTGGAGCATTCCGCTCACTGTTTCGGAGGTA 2299

Qy 1381 TGTCTTGGATCACACAGGGGCTTCTGGGAGCTCTTCTGCTGTGGATGGGAATTAACGCCC 1440
 ||||| || ||||| | ||||| ||||| ||||| ||||| || || || |
 Db 2300 TGTCTTGGATAACGCAGGGATTGCTGGGGGCTCTTCTGTTGTGGATGGGCATCAATGCTC 2359

Qy 1441 GTGACAGGTCAATTGCTATGACGTTCTTGGCGTTGGAGGAGTCTTGTCTTCTCTTTCGG 1500
 ||||| || || | ||||| || || ||||| ||||| ||||| || || |
 Db 2360 GTGACAGGTCCATAGCTCTCACGTTTCTCGCAGTTGGAGGAGTCTTGTCTTCTCTCTG 2419

Qy 1501 TCAACGTCCATGCTG 1515
 | ||||| || |||||
 Db 2420 TGAACGTGCACGCTG 2434

RESULT 11

AF404757

LOCUS AF404757 11029 bp ss-RNA linear VRL 23-JUL-2002

DEFINITION West Nile virus isolate WN Italy 1998-equine, complete genome.

ACCESSION AF404757

VERSION AF404757.1 GI:21929240

KEYWORDS .

SOURCE West Nile virus

ORGANISM West Nile virus

Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 Flavivirus; Japanese encephalitis virus group.

REFERENCE 1 (bases 1 to 11029)

AUTHORS Lanciotti,R.S., Ebel,G.D., Deubel,V., Kerst,A.J., Murri,S.,
 Meyer,R., Bowen,M., McKinney,N., Morrill,W.E., Crabtree,M.B.,
 Kramer,L.D. and Roehrig,J.T.

TITLE Complete genome sequences and phylogenetic analysis of West Nile
 virus strains isolated from the United States, Europe, and the
 Middle East

JOURNAL Virology 298 (1), 96-105 (2002)

PUBMED 12093177

REFERENCE 2 (bases 1 to 11029)

AUTHORS Deubel,V., Bowen,M., Meyer,R., McKinney,N. and Morrill,W.

TITLE Direct Submission

JOURNAL Submitted (02-AUG-2001) Division of Vector-Borne Infectious
 Diseases, Centers for Disease Control & Prevention, Rampart Road,
 Fort Collins, CO 80521, USA

FEATURES Location/Qualifiers

source 1. .11029
 /organism="West Nile virus"
 /mol_type="genomic RNA"
 /isolate="WN Italy 1998-equine"
 /specific_host="equine"
 /db_xref="taxon:11082"
 /country="Italy"

CDS 97. .10398

/note="contains capsid, pre-membrane, envelope, NS1, NS2a, NS2b, NS3, NS4a, NS4b, and NS5"
/codon_start=1
/product="polyprotein precursor"
/protein_id="AAM81753.1"
/db_xref="GI:21929241"
/translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGP
RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLISFKKELGTLTSAINRRSSKQ
KKRGGKTGIAVMIGLIASVGAUTLSNFQGVMMTNATDVTDVITPTAAGKNLCIVR
AMDVGVMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
VFVVLILLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCTVIMSKDKPTIDVK
MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRAPAFVCRQGVVDRGWG
NGCGLFGKGSIDTCAKFACSTKATGRTILKENIKYEVAIFVHGPTTVESHGNYSTQIG
ATQAGRFSITPAAPSYTLKLGEYGEVTVDCPRSGIDTNAYVMTVGTKTFLVHREWF
MDLNLPSWSSAGSTVWRNRETLMFEFEPHATKQSVIALGSQEGALHQALAGAI PVEFSS
NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
PCKVPISSVASLNDLTPVGRLVTVPFVSVATANAKVLEIEPPFGDSYIVVGRGEQQ
INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSGGVFTSVGKAVHQVFGGAF
RSLFGGMSWITQGLLGLALLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
ISRQELRCGSGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
QMWESVKDELNTLLKENGVDLSVVVEKQEGMYKSAPKRLTATTEKLEIGKAWGKSIL
FAPELANNTFVVDGPETKECPTQNRANSLEVEDFGFGLTSTRMFLKVRESNTTECD
KIIIGTAVKNNLAIHSDLSYWIESRLNDTWKLERAVLGEVKSCTWPETHTLWGDGILES
DLIIPVTLAGPRSNHNRRPGYKTQNGQPWDEGRVEIDFDYCPGTTVTLSSECGHRGPA
TRTTTESGKLITDWCCRSCTLPPLRYQTDSCGWYGMIEIRPQRHDEKITYLVQSVNAYNA
DMIDPFQLGLLVFLATQEVLRKRWTAKISMPAILIALLVLFVGGITYTDVLRVILV
GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILMLAAVFFQMAYH
DARQILLWEIPDVLNSLAVAWMILRAITFTTTSNVVPLALLTPGLRCLNLDVYRIL
LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRGWPA
TEVMTAVGLMFAIVGGLAELDIDSMAPMTIAGLMFAAFVISGKSTDMWIERTADISW
ESDAEITGSSSERVDVRLDDGNFQLMNDPGAPWKIWMRLMACLAISAYTPWAILPSVI
GFWITLQYTKRGGVLWDTPSPKEYKKGDTTGVYRIMTRGLLSYQAGAGVMVEGVFH
TLWHTTKGAALMSGEGRLDPYWGSKEDRLCYGGPWKLQHKWNGQDEVQMI VVEPGKN
VKNVQTKPGVFKTPEGEIGAVTLDFTGTSGSPIVDKNGDVIGLYNGVIMPNGSYIS
AIVQGERMDEPI PAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
LAPTRVVAEAMAEALRGLPIRYQTSAVTREHNGNEIVDMCHATLTHRLMSPHRVPNY
NLFVMDAEHFTDPASIAARGYISTKVELGEAAAFMTATPPGTSDFPFESNSPISDLQ
TEIPDRAWNSGYEWITEYIGKTWVFPVSVKMGNEIALCLQRAGKKVVQLNRKSYET
PKCKNDWDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEP
SAVTAASAAQRRGRIGRNPQVGEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLIAQFYQ
PEREKVYTMDEYRLRGEERKNFLELLRTADLPVWLAYKVAAGVSYHRRWCDFDGP
RNTILEDNNEVEVITKLGERKILRPWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHMALEELPDALQTIALLSVMTMG
VFFLLMQRKIGIKIGLGGVVLGVATFFCWMAEVPGTKIAGMLLLSLLMIVLPIPEK
QRSQTDNQLAVFLICVMTLVSAVAANEMGWLDKTKSDISSLFQRIEVKENFSMGEFL
LDLRPATAWSLYAVTTAVLTPLLKHLITSYINTSLTSINVQASALFTLARGFPFVDV
GVSALLLAAGCWGQVTLTVTVAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKN
VVDGIVATDVPELERTTPIMQKKVGQIMLILVSLAAVVVNPSVKTVREAGILITAAV
TLWENGASSVWNATTAIGLCHIMRGGWLSCLSTWTLIKNMDKPLKRGGA
KGRTLGEVWKERLNQMTKEEFTRYRKEAIEVDRSAAKHARKEGNTVGGHPVSRGTAKLRWLVER
RFLEPVGKVIDLGCGRGGWCYMATQKRVQEVRYTKGGPGHEEPQLVQSYGWNIVTM
KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRRGPREFCVKVL
CPYMPKVI EKMELLQRRYGGGLVRNPLSRNSTHEMYWVSASGNVVSVMNTSQVLLG
RMEKRTWKGPQYEEDVNLGSGTRAVGKPLNSDTSKIKNRIERLRREYSSTWHHDENH
PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDITINVTTMAMTDTTPFGQQRVFK
EKVDTKAPEPEGVKYVLNETTNWLWAFALAREKPRMCSREEFIRKVNSNAALGAMFE

Db	1556	ACACTAATGCCCTACTACGTGATGACTGTTGGAAACAAAGACGTTTTTTGGTCCATCGTGAGT	1615
Qy	661	GGTTTATGGACCTGAACCTTCCATGGAGTAGCGCTGGAAGCACAAACGTGGAGGAACCGGG	720
Db	1616	GGTTCATGGACCTCAACCTCCCTTGGAGCAGTGCTGGAAGTACTGTGTGGAGGAACAGAG	1675
Qy	721	AAACACTGATGGAGTTTGAAGAACCTCATGCCACCAAACAATCTGTCTAGCTCTAGGGT	780
Db	1676	AGACGTTGATGGAGTTTGAAGAACACACGCCACAAAGCAGTCTGTGATAGCATTGGGGCT	1735
Qy	781	CGCAGGAAGGTGCCTTGCACCAAGCTCTGGCTGGAGCAATTCTGTGAGTTCTCAAGCA	840
Db	1736	CACAAGAGGGAGCTCTGCATCAAGCTTTGGCTGGAGCCATCCCTGTGGAATTTTCAAGCA	1795
Qy	841	ACACTGTGAAGTTGACATCAGGACATCTGAAGTGTAAGGTGAAGATGGAGAAGTTGCAGC	900
Db	1796	ACACTGTCAAGTTGACGTCGGGCCATTTGAAGTGTAAGTGAAGATGGAAAAATTGCAGT	1855
Qy	901	TGAAGGGAACAACATATGGTGTATGCTCAAAAGCATTCAAATTCGCTAGGACTCCCGCTG	960
Db	1856	TGAAGGGAACAACCTTACGGCGTCTGTTCAAAGGCTTTCAAGTTTCTTGGGACTCCCGCAG	1915
Qy	961	ACACTGGTCATGGAACGGTGGTGCTGGAAGTGCAGTATACCGGAAAAGACGGGCCTTGCA	1020
Db	1916	ACACAGGTCACGGCACTGTGGTGTGGAATTGCAGTACACTGGCACGGATGGACCTTGCA	1975
Qy	1021	AAGTGCCCATTTCTTCTGTGGCTTCCCTGAACGACCTTACACCCGTTGGAAGGCTGGTGA	1080
Db	1976	AAGTTCCCATCTCGTCAGTGGCTTCATTGAACGACCTAACACCCGGTAGGCAGATTGGTCA	2035
Qy	1081	CTGTGAATCCATTTGTGTCTGTGGCTACGGCCAACCTCGAAGGTTTTGATTGAACTCGAAC	1140
Db	2036	CTGTTAACCCTTTTGTTCAGTGGCCACGGCCAACGCCAAGGTCCTGATTGAACTGGAAC	2095
Qy	1141	CCCCGTTTAGTGACTCTTACATCGTGGTGGGGAGAGGAGAACAGCAGATAAACACCACCT	1200
Db	2096	CACCCCTTGGAGACTCATACATAGTGGTAGGCAGAGGAGAACAACAGATCAATCACCATT	2155
Qy	1201	GGCACAAATCTGGGAGCAGTATTGGAAGGCTTTCACCACTACACTCAGAGGAGCTCAAC	1260
Db	2156	GGCATAAGTCTGGAAGCAGCATTGGCAAAGCCTTTACAACCACTCTCAAAGGGGCGCAGA	2215
Qy	1261	GACTTGCAGCTCTTGGAGACACTGCCTGGGATTTTGGATCAGTCGGAGGGGTTTTACCT	1320
Db	2216	GATTAGCCGCTCTAGGAGACACAGCTTGGGACTTCGGATCAGTTGGAGGGGTGTTTACCT	2275
Qy	1321	CGGTAGGGAAAGCCATACACCAAGTTTTTTGGAGGAGCCTTTAGATCACTCTTTGGAGGGA	1380
Db	2276	CAGTAGGGAAGGCTGTCCATCAAGTGTTCCGGTGGAGCAATCCGCTCACTGTTCCGAGGTA	2335
Qy	1381	TGTCCTGGATCACACAGGGGCTTCTGGGAGCTCTTCTGCTGTGGATGGGAATTAACGCCC	1440
Db	2336	TGTCCTGGATAACGCAGGGATTGCTGGGGGCTCTTCTGTTGTGGATGGGCATCAATGCTC	2395
Qy	1441	GTGACAGGTCAATTGCTATGACGTTCCCTTGCGGTTGGAGGAGTCTTGCTCTTCCCTTTCGG	1500
Db	2396	GTGACAGGTCCATAGCTCTCACGTTTCTCGCAGTTGGAGGAGTCCTGCTCTTCCCTCTCCG	2455

Qy 1501 TCAACGTCCATGCTG 1515
| | | | | | | | | |
Db 2456 TGAACGTGCACGCTG 2470

RESULT 12

AY660002

LOCUS AY660002 11029 bp RNA linear VRL 19-DEC-2004

DEFINITION West Nile virus isolate Mex03 from Mexico, complete genome.

ACCESSION AY660002

VERSION AY660002.1 GI:55975602

KEYWORDS .

SOURCE West Nile virus (WNV)

ORGANISM West Nile virus

Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Flavivirus; Japanese encephalitis virus group.

REFERENCE 1 (bases 1 to 11029)

AUTHORS Beasley,D.W.C., Davis,C.T., Estrada-Franco,J., Navarro-Lopez,R.,
Campomanes-Cortes,A., Tesh,R.B., Weaver,S.C. and Barrett,A.D.T.

TITLE Genome Sequence and Attenuating Mutations in West Nile Virus
Isolate from Mexico

JOURNAL Emerging Infect. Dis. 10 (12), 2221-2224 (2004)

REFERENCE 2 (bases 1 to 11029)

AUTHORS Beasley,D.W.C., Davis,T., Estrada-Franco,J.G., Tesh,R.B.,
Weaver,S.C. and Barrett,A.D.T.

TITLE Direct Submission

JOURNAL Submitted (18-JUN-2004) Pathology, University of Texas Medical
Branch, 301 University Blvd., Galveston, TX 77555-0609, USA

FEATURES Location/Qualifiers

source

1. .11029

/organism="West Nile virus"

/mol_type="genomic RNA"

/strain="TM171-03"

/isolate="Mex03"

/db_xref="taxon:11082"

/country="Mexico"

CDS

97. .10398

/codon_start=1

/product="polyprotein precursor"

/protein_id="AAV68177.1"

/db_xref="GI:55975603"

/translation="MSKKPGGPGKSRVNMMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
KKRGGKTGIAVMIGLIASVGAVTLSNFQGVMMTVNATDVTDVITPTAAGKNLCIVR
AMDVGVMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVTGWMLGSNTMQRV
VFVVLILLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSVITMSKDKPTIDVK
MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRAPAFVCRQGVVDRGWG
NGCGLFGKGSIDTCAKFACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYPTQVG
ATQAGRFSITPAAPSYTLKLGEYGEVTVDCPRSGIDTNAYYVMTVGTKTFLVHREWF
MDLNLPSWSSAGSTVWRNRETLMFEFEPHATKQSVIALGSQEGALHQALAGAI PVEFSS
NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
PCKVPISSVASLNDLTPVGRLVTNPFVSVATANAKVLIIELEPPFGDSYIVVGRGEQQ
INHHWHKSGSSIGKAFTTTTLKGAQRLAALGDTAWDFGSGGVFTSVGKAVHQVFVGGAF
RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
ISRQELRCGSGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH

QMWEAVKDELNTLLKENGVDLSVVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGKSIL
 FAPELANNTFVVDGPETKECPTQNRANSLEVEDFGFGLTSTRMFLKVRESNTTECDSDS
 KIIGTAVKNNLAHSDLSYWIESRLNDTWKLERAVLGEVKSCTWPETHTLWGDGILES
 DLIIPVTLAGPRSNHNRPGYKTQNGQPWDEGRVEIDFDYCPGTTVTLSSECGHRGPA
 TRTTTESGKLITDWCCRSTLPPLRYQTDSCGWYGMIEIRPQRHDEKTLVQSQVNAYNA
 DMIDPFQGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDLVRYVILV
 GAAFAESNSGGDVVHLALMATFKIQPFVMVASFLKARWTNQENILLMLAAVFFQMAYH
 DARQILLWEIPDVLNSLAVAWMILRAITFTTTSNVVPLLALLTPGLRCLNLDVYRIL
 LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMLAAGLIACDPNRKRGWPA
 TEVMTAVGLMFAIVGGLAELDISMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
 ESDAEITGSSERVDVRLDDDNFQLMNDPGAPWKIWMRLRMVCLAISAYTPWAILPSVV
 GFWITLQYTKRGGVLWDTPSPKEYKKGDTTGTGVYRIMTRGLLGSYQAGAGVMVEGVFH
 TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
 VKNVQTKPGVFKTPEGEIGAVTLDFTGTSGSPIVDKNGDVIGLYNGVIMPNGSYIS
 AIVQGERMDEPI PAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
 LAPTRVVAEAEALRGLPIRYQTSAPVREHNGNEIVDMCHATLTHRLMSPHRVPNY
 NLFVMDAEAHFTDPASIAARGYISTKVELGEAAAFMTATPPGTSDFPFESNSPISDLQ
 TEIPDRAWNSGYEWITEYTGKTWVFVPSVKMGNEIALCLQRAGKKVVQLNRKSYET
 PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEP
 SAVTAA SAAQRRGRIGRNP SQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLIAQFYQ
 PEREKVYTMDEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVSYHRRWCFDGP
 R TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
 LGKMPHEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALLSVMTMG
 VFFLLMQRKIGIGLGGAVLG VATFFCWM AEVPGTKIAGMLLLSLLMIVLIP EPEK
 QRSQTDNQLAVFLICVMTLVSAVANEMGWLDKTKSDISSLFGQRIEVKENFSMGEFL
 LDLRPATAWSLYAVTTAVLTPLLKLHITSYINTSLTSINVQASALFTLARGFPFVDV
 GVSALLLAAGCWGQVTLTVTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
 VVDGIVATDVPELERTTPIMQKKVQIMLILVSLAAVVVNP SVKTVREAGILITAAAV
 TLWENGASSVWNATTAIGLCHIMRGGWLSCL SITWTLVKNMEKPG LKRGGA KGR TLGE
 VWKERLNQMTKEEFTRYRKEAII EVDRSAAKHARKEGNTVGGHPVSRGTAKLRWLVER
 RFLEPVGKVIDLGCGRGGWCYMATQKRVQEVRYTKGGPGHEEPQLVQSYGWNIVTM
 KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFCVKVL
 CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSASGNVSVHVNMTSQVLLG
 RMEKRTWKGPQYEEDVNLGSGTRAVGKPLNLSDTSKIKNRIERLRREYSSTWHHDENH
 PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDITITNVTTMAMTDTTPFGQQRVFK
 EKVDTKAPEPEGVKYVLNETTNWLWAF LAREKRPRMCSREEFIRKVNSNAALGAMFE
 EQNQWRSAREAVEDPKFWEMVDEEREAH LRGECHTCIYNMMGKREKKPGEFGKAKGSR
 AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGT
 RPPGGKI YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAI IELTYRHKVVKMRPAADGR
 TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTGKGPKV
 RTWLFENGEEERLSRMAVSGDDCVVKPLDDR FATS LHLFNAMSKVRKDIQEWK
 PSTGWY DWQQVPFCSNHFTELIMKDGR TLVVP CRGQDELVGRARISP
 GAGWNVRDTACLA KSYA QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGR
 TTWSIHAGGEWMTTEDMLEVWNRVW IEENEWMEDKTPVEKWS
 DVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRAIIG
 DEKYVDYMSSLKRYEDTILVEDTVL"

mat_peptide	97. .465	/product="nucleocapsid protein"
		/note="C"
mat_peptide	466. .741	/product="pre-membrane protein"
		/note="prM"
mat_peptide	742. .966	/product="membrane protein"
		/note="M"
mat_peptide	967. .2469	/product="envelope glycoprotein"
		/note="E"

Db 1376 ACGAAGTGGCCATTTTTGTCCATGGACCAACTACTGTGGAGTCGCACGGAAACTACCCCA 1435
 Qy 481 CACAGATAGGGGCTACCCAAGCAGGAAGGTTTACGCATAACTCCATCGGCACCATCCTACA 540
 ||||| | || || || || ||||| || ||||| ||||| ||||| || || |||||
 Db 1436 CACAGGTTGGAGCCACTCAGGCAGGGAGATTACGCATCACTCCTGCGGCGCCTTCATACA 1495
 Qy 541 CGCTGAAGTTGGGTGAGTATGGTGAGGTCACAGTTGACTGTGAGCCACGGTCAGGAATAG 600
 | || ||| | || || ||||| ||||| ||||| ||||| ||||| ||||| || ||
 Db 1496 CACTAAAGCTTGGAGAATATGGAGAGGTGACAGTGGACTGTGAACCACGGTCAGGGATTG 1555
 Qy 601 ACACTAGCGCTTACTACGTTATGTGAGTGGGTGCGAAGTCCTTCTTGGTTTACCCGAGAAT 660
 ||||| | || ||||| ||||| || || || || || || || ||||| || || || ||
 Db 1556 ACACCAATGCATACTACGTGATGACTGTTGGAACAAAGACGTTCTTGGTCCATCGTGAGT 1615
 Qy 661 GGTTTATGGACCTGAACCTTCCATGGAGTAGCGCTGGAAGCACACGTGGAGGAACCGGG 720
 ||||| ||||| || ||||| || ||||| || ||||| || ||||| ||||| || ||
 Db 1616 GGTTTATGGATCTCAACCTCCCTTGGAGCAGTGCTGGAAGTACTGTGTGGAGGAACAGAG 1675
 Qy 721 AAACACTGATGGAGTTTGAAGAACCTCATGCCACCAACAATCTGTGCTAGCTCTAGGGT 780
 | || | ||||| ||||| ||||| || ||||| || ||||| ||||| || || ||
 Db 1676 AGACGTTAATGGAGTTTGAAGAACACACGCCACGAAGCAGTCTGTGATAGCATTGGGCT 1735
 Qy 781 CGCAGGAAGGTGCCTTGACCAAGCTCTGGCTGGAGCAATTCCTGTTGAGTTCTCAAGCA 840
 | || || || || || ||||| ||||| ||||| ||||| ||||| || || |||||
 Db 1736 CACAAGAGGGAGCTCTGCATCAAGCTTTGGCTGGAGCCATTCTGTGGAATTTTCAAGCA 1795
 Qy 841 ACACTGTGAAGTTGACATCAGGACATCTGAAGTGTAGGGTGAAGATGGAGAAGTTGCAGC 900
 ||||| ||||| || || || || || ||||| ||||| ||||| ||||| || |||||
 Db 1796 ACACTGTCAAGTTGACGTCGGGTCATTTGAAGTGTAGAGTGAAGATGGAAAATTGCAGT 1855
 Qy 901 TGAAGGGAACAACATATGGTGTATGCTCAAAAGCATTCAAATTCGCTAGGACTCCCCTG 960
 ||||| ||||| ||||| || || || || || ||||| || ||||| || ||||| ||
 Db 1856 TGAAGGGAACAACCTATGGCGTCTGTTCAAAGGCTTTCAAGTTTCTTGGGACTCCCGCAG 1915
 Qy 961 ACACTGGTCATGGAACGGTGGTGCTGGAAGTGCAGTATACCGGAAAAGACGGGCCTTGCA 1020
 ||||| ||||| || || ||||| ||||| ||||| || || || || || || |||||
 Db 1916 ACACAGGTCACGGCACTGTGGTGTGGAATTGCAGTACACTGGCACGGATGGACCTTGCA 1975
 Qy 1021 AAGTGCCCATTTCTTCTGTGGCTTCCCTGAACGACCTTACACCCGTTGGAAGGCTGGTGA 1080
 ||||| || || || || || ||||| ||||| ||||| || || || || || || |||||
 Db 1976 AAGTTCCTATCTCGTCAGTGGCTTCATTGAACGACCTAACGCCAGTGGGCAGATTGGTCA 2035
 Qy 1081 CTGTGAATCCATTTGTGTCTGTGGCTACGGCCAACCTCGAAGGTTTTGATTGAACTCGAAC 1140
 ||||| || || ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 2036 CTGTCAACCCTTTTGTTTCAGTGGCCACGGCCAACGCTAAGGTCCTGATTGAATTGGAAC 2095
 Qy 1141 CCCCCTTTAGTGACTCTTACATCGTGGTGGGGAGAGGAGAACAGCAGATAAACCACCACT 1200
 | || ||| | ||||| ||||| ||||| ||||| ||||| ||||| ||||| || |||||
 Db 2096 CACCCTTTGGAGACTCATACATAGTGGTGGGCAGAGGAGAACAACAGATCAATCACCATT 2155
 Qy 1201 GGCACAAATCTGGGAGCAGTATTGGAAAGGCTTTTACCACCTACACTCAGAGGAGCTCAAC 1260
 ||||| ||||| ||||| ||||| || || || || || || || ||||| ||||| ||
 Db 2156 GGCACAAGTCTGGAAGCAGCATTGGCAAAGCCTTTACAACCACCCTCAAAGGAGCGCAGA 2215
 Qy 1261 GACTTGCAGCTCTTGGAGACACTGCCTGGGATTTTGGATCAGTCGGAGGGGTTTTACCT 1320
 ||||| || ||||| ||||| || ||||| ||||| ||||| ||||| ||||| |||||
 Db 2216 GACTAGCCGCTCTAGGAGACACAGCTTGGGACTTTGGATCAGTTGGAGGGGTGTTACCT 2275

Qy 1321 CGGTAGGGAAAGCCATACACCAAGTTTTTGGAGGAGCCTTTAGATCACTCTTTGGAGGGA 1380
 | || |||| | | | |||| | | |||| | | |||| | | |||| |
 Db 2276 CAGTTGGGAAGGCTGTCCATCAAGTGTTCGGAGGAGCATTCGCTCACTGTTTGGAGGCA 2335
 Qy 1381 TGTCTGGATCACACAGGGGCTTCTGGGAGCTCTTCTGCTGTGGATGGGAATTAACGCCC 1440
 |||| |||| | | || | |||| |||| | | |||| |||| | | ||
 Db 2336 TGTCTGGATAACGCAAGGATTGCTGGGGGCTCTCCTGTTGTGGATGGGCATTAATGCTC 2395
 Qy 1441 GTGACAGGTCAATTGCTATGACGTTCCCTTGC GGTTGGAGGAGTCTTGCTCTTCCTTTCGG 1500
 |||| |||| | | || | |||| || || || |||| |||| | | ||
 Db 2396 GTGATAGGTCCATAGCTCTCACGTTTCTCGCAGTTGGAGGAGTTCTGCTCTTCCTCTCCG 2455
 Qy 1501 TCAACGTCCATGCTG 1515
 | |||| | ||||
 Db 2456 TGAACGTGCATGCTG 2470

RESULT 13

DQ164190

LOCUS DQ164190 11029 bp RNA linear VRL 18-NOV-2005

DEFINITION West Nile virus isolate NY 2003 Suffolk, complete genome.

ACCESSION DQ164190

VERSION DQ164190.1 GI:76781539

KEYWORDS .

SOURCE West Nile virus (WNV)

ORGANISM West Nile virus

Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Flavivirus; Japanese encephalitis virus group.

REFERENCE 1 (bases 1 to 11029)

AUTHORS Davis,C.T., Ebel,G.D., Lanciotti,R.S., Brault,A.C., Guzman,H., Siirin,M., Lambert,A., Parsons,R.E., Beasley,D.W., Novak,R.J., Elizondo-Quiroga,D., Green,E.N., Young,D.S., Stark,L.M., Drebot,M.A., Artsob,H., Tesh,R.B., Kramer,L.D. and Barrett,A.D.

TITLE Phylogenetic analysis of North American West Nile virus isolates, 2001-2004: Evidence for the emergence of a dominant genotype

JOURNAL Virology 342 (2), 252-265 (2005)

PUBMED 16137736

REFERENCE 2 (bases 1 to 11029)

AUTHORS Davis,T.C., Ebel,G.D., Lanciotti,R.S. and Brault,A.C.

TITLE Direct Submission

JOURNAL Submitted (11-AUG-2005) Pathology, University of Texas Medical Branch, 301 University Blvd., Galveston, TX 77550, USA

FEATURES Location/Qualifiers

source 1. .11029
 /organism="West Nile virus"
 /mol_type="genomic RNA"
 /isolate="NY 2003 Suffolk"
 /specific_host="American crow"
 /db_xref="taxon:11082"
 /country="USA: NY, Suffolk"
 /collection_date="2003"
 CDS 97. .10398
 /note="encodes C, prM, E, NS1, NS2A, NS2B, NS3, NS4A, NS4B, NS5"
 /codon_start=1
 /product="polyprotein precursor"

/protein_id="ABA54579.1"

/db_xref="GI:76781540"

/translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
RFVLALLAFFRFTAIAPTRAVLNRRWGVNKQTAMKHLISFKKELGTLTSAINRRSSKQ
KKRGGKTGIAVMIGLIASVGAVTLSNFQGVMMTVNATDVTDVITPTAAGKNLCIVR
AMDVGVMCDDTITYECPVLSAGNDPEDIDCWCTKSAYVRYGRCTKTRHSRRSRRSLT
VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
VFVVLVLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
MMNMEAANLAEVRSYCYLATVSDLSKAACPTMGEAHNDKRAPAFVCRQGVVDRGWG
NGCGLFGKGSIDTCAKFACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYSTQAG
ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
MDLNLPSWSSAGSTVWRNRRETLMFEFEPHATKQSVIALGSGQEGALHQALAGAIPEVFSS
NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
PCKVPISSVASLNDLTPVGRVLTVPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQ
INHHWHKSGSSIGKAFTTTTLKGAQRLAALGDTAWDFGSGVGGVFTSVGKAVHQVFGGAF
RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
ISRQELRCGSGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
QMWEAVKDELNTLLKENGVDLSVVVEKQEGMYKSAPKRLTATTEKLEIGKAWGKSIL
FAPELANNTFVVDGPETKECPTQNRAWSLEVEDFGFGLTSTRMFLKVRESNTTECDS
KII GTAVKNNLAIHSDLSYWIESRLNDTWKLERAVLGEVKSCTWPETHTLWGDGILES
DLIIPVTLAGPRSNHNRRPGYKTQNGQGPWDEGRVEIDFDYCPGTTVTLSSESCGHRGPA
TRTTTESGKLITDWCCRSCITLPLRYQTDSGCWYGMEIRPQRHDEKTLVQSQVNAYNA
DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRVILV
GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILMLAAVFFQMAYH
DARQILLWEIPDVLSLAVAWMILRAITFTTTSNVVPLLALLTPGLRCLNLDVYRIL
LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRGWPA
TEVMTAVGLMFAIVGGLAELDISMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
ESDAEITGSSSERVDVRLDDDDGNFQLMNDPGAPWKIWMRLMVC LAISAYTPWAILPSVV
GFWITLQYTKRGGVLWDTSPSPKEYKKGDTTGTGVYRIMTRGLLGSYQAGAGVMVEGVFH
TLWHTTKGAALMSGEGRLDPYWGVSVEDRLCYGGPWKLQHKWNGQDEVQMI VVEPGKN
VKNVQTKPGVFKTPEGEIGAVTLDFTGTSGSPIVDKNGDVIGLYGNVIMPNGSYIS
AIVQGERMDEPI PAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
LAPTRVVAEMAELRGLPIRYQTSAPPREHNGNEIVDMCHATLTHRLMSPHRVPNY
NLFVMDAEHFTDPASIAARGYISTKVELGEAAAI FMTATPPGTSDFPFESNSPISDLQ
TEIPDRAWNSGYEWITEYTGKTWVFPVSVKMGNEIALCLQRAGKKVVQLNRKSYET
PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEP
SAVTAASAAQRRGRIGRNP SQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLIAQFYQ
PEREKVYTM DGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVSYHRRWC
FDGPR TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
LGKMP EHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALLIALLSVMTMG
VFFLLMQRKIGIGKIGLGGAVLG VATFFCWMAEVPGTKIAGMLLLSLLMIVLIP
EPEK QRSQTDNQLAVFLICVMTLVSAVANEMGWLDKTKSDISSLFGQRIE
VKENFSMGEFL LDLRPATAWSLYAVTTAVLTPLLKHLITS
DYINTSLTSINVQASALFTLARGFPFVDV GVSALLAAGCWGQVTLTVT
VTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA VVDGIVATDVPELERT
TPIMQKKVGQIMLILVSLAAVVVNPSVKTVREAGILITAAAV TLWENGASSV
WNATTAIGLCHIMRGGWLSCLSITWTLIK NMEKPG LKRGGA
KGRTLGE VWKERLNQMTKEEFTRYRKEAII EVDRSAAKHARKEGNVTGGHPV
SRGTAKLRWLVER RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEV
RGYTKGGPGHEEPQLVQS YGWNIVTM KSGVDVFYRPSECCD
TLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFCVKVL
CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVS
RASGNVHVSVMNTSQVLLG RMEKRTWKGPQYEEDVNLGSG
TRAVGKPLNSDTRKIKNRIERLRREYSSTWHHDENH
PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKP
WDTITNVTTMAMTDTTPFGQQRVFK EKVDTKAPEP
EGVKYVLNETTNWLWAF LAREKRPRMCSREEFI
RKVNSNAALGAMFE EQNQWRSAREAVEDPKF
WEMVDEERE AHLRGECHTCIYNMMGKREKKP
GEFGKAKGSR AIWFMWLGARFLEFEALGFLNED
HWLGRKNSGGGVEGLGLQKLG YILREVGTRP
GGKI YADDTAGWDTRITRADLENEAKVLELLDGE
HRRRLARAI IELTYRHKVVKVMRPAADGR
TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVR
MMEGEGVIGPDDVEKLTGKGKGPKV

Db	1616	GGTTCATGGACCTCAACCTCCCTTGGAGCAGTGTCTGGAAGTACTGTGTGGAGGAACAGAG	1675
Qy	721	AAACACTGATGGAGTTTGAAGAACCTCATGCCACCAACAATCTGTCTAGCTCTAGGGT	780
Db	1676	AGACGTTAATGGAGTTTGAGGAACCACACGCCACGAAGCAGTCTGTGATAGCATTGGGCT	1735
Qy	781	CGCAGGAAGGTGCCTTGCACCAAGCTCTGGCTGGAGCAATTCCTGTTGAGTTCTCAAGCA	840
Db	1736	CACAAGAGGGAGCTCTGCATCAAGCTTTGGCTGGAGCCATTCTGTGGAATTCTCAAGCA	1795
Qy	841	ACACTGTGAAGTTGACATCAGGACATCTGAAGTGTAGGGTGAAGATGGAGAAGTTGCAGC	900
Db	1796	ACACTGTCAAGTTGACGTCGGGTCAATTTGAAGTGTAGAGTGAAGATGGAAAAATTGCAGT	1855
Qy	901	TGAAGGGAACAACATATGGTGTATGCTCAAAAGCATTCAAATTCGCTAGGACTCCCCTG	960
Db	1856	TGAAGGGAACAACCTATGGCGTCTGTTCAAAGGCTTTCAGTTTCTTGGGACTCCCGCAG	1915
Qy	961	ACACTGGTCATGGAACGGTGGTGTGGAAGTGCAGTATACCGGAAAAGACGGGCCTTGCA	1020
Db	1916	ACACAGGTCACGGCACTGTGGTGTGGAATTGCAGTACACTGGCACGGATGGACCTTGCA	1975
Qy	1021	AAGTGCCCATTTCTTCTGTGGCTTCCCTGAACGACCTTACACCCGTTGGAAGGCTGGTGA	1080
Db	1976	AAGTTCCTATCTCGTCAGTGGCTTCATTGAACGACCTAACGCCAGTGGGCAGATTGGTCA	2035
Qy	1081	CTGTGAATCCATTTGTGTCTGTGGCTACGGCCAACTCGAAGGTTTTGATTGAACTCGAAC	1140
Db	2036	CTGTCAACCCTTTTGTTCAGTGGCCACGGCCAAACGCTAAGGTCCTGATTGAATTGGAAC	2095
Qy	1141	CCCCGTTTAGTGACTCTTACATCGTGGTGGGGAGAGGAGAACAGCAGATAAACCACCACT	1200
Db	2096	CACCTTTTGGAGACTCATACATAGTGGTGGGCAGAGGAGAACAACAGATCAATCACCATT	2155
Qy	1201	GGCACAAATCTGGGAGCAGTATTGGAAAGGCTTTCACCACTACACTCAGAGGAGCTCAAC	1260
Db	2156	GGCACAAAGTCTGGAAGCAGCATTTGGCAAAGCCTTTACAACCACCCTCAAAGGAGCGCAGA	2215
Qy	1261	GACTTGCAGCTCTTGAGACACTGCCTGGGATTTTGGATCAGTCGGAGGGGTTTTACCT	1320
Db	2216	GACTAGCCGCTCTAGGAGACACAGCTTGGGACTTTGGATCAGTTGAGGGGTGTTACCT	2275
Qy	1321	CGGTAGGGAAAGCCATACACCAAGTTTTTGGAGGAGCCTTTAGATCACTCTTTGGAGGGA	1380
Db	2276	CAGTTGGGAAGGCTGTCCATCAAGTGTTCGGAGGAGCATTCGCTCACTGTTTCGGAGGCA	2335
Qy	1381	TGTCCTGGATCACACAGGGGCTTCTGGGAGCTCTTCTGCTGTGGATGGGAATTAACGCCC	1440
Db	2336	TGTCCTGGATAACGCAAGGATTGCTGGGGGCTCTCCTGTTGTGGATGGGCATCAATGCTC	2395
Qy	1441	GTGACAGGTCAATTGCTATGACGTTCCCTTGCGGTTGGAGGAGTCTTGCTCTTCTTTCCG	1500
Db	2396	GTGATAGGTCCATAGCTCTCACGTTTCTCGAGTTGGAGGAGTCTGCTCTTCTCTCCG	2455
Qy	1501	TCAACGTCCATGCTG	1515
Db	2456	TGAACGTGCATGCTG	2470

RESULT 14
AY371271
LOCUS AY371271 2004 bp RNA linear VRL 25-NOV-2003
DEFINITION West Nile virus strain TM171-03 polyprotein gene, partial cds.
ACCESSION AY371271
VERSION AY371271.1 GI:38224786
KEYWORDS .
SOURCE West Nile virus (WNV)
ORGANISM West Nile virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Flavivirus; Japanese encephalitis virus group.
REFERENCE 1 (bases 1 to 2004)
AUTHORS Estrada-Franco, J.G., Navarro-Lopez, R., Beasley, D.W.C., Coffey, L.,
Carrara, A.-S., Travassos da Rosa, A., Clements, T., Wang, E.,
Ludwig, G.V., Campomanes Cortes, A., Paz Ramirez, P., Tesh, R.B.,
Barrett, A.D.T. and Weaver, S.C.
TITLE West Nile virus in Mexico: evidence of widespread circulation since
July, 2002
JOURNAL Emerging Infect. Dis. 9 (12), 1604-1607 (2003)
REFERENCE 2 (bases 1 to 2004)
AUTHORS Beasley, D.W.C., Estrada-Franco, J.G., Tesh, R.B., Weaver, S.C. and
Barrett, A.D.T.
TITLE Direct Submission
JOURNAL Submitted (20-AUG-2003) Pathology, University of Texas Medical
Branch, 301 University Blvd., Galveston, TX 77555-0609, USA
FEATURES
source Location/Qualifiers
1. .2004
/organism="West Nile virus"
/mol_type="genomic RNA"
/strain="TM171-03"
/db_xref="taxon:11082"
/country="Mexico"
CDS
<1. .>2004
/codon_start=1
/product="polyprotein"
/protein_id="AAR14153.1"
/db_xref="GI:38224787"
/translation="VTLSNFQGKVMNTVNATDVTDVITPTAAGKNLCIVRAMDVGYM
CDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLTVQTHGES
TLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVTGWM LGSNTMQRVVFVLLL
LVAPAYSFNC LGMSNRDFLEGVSGATWVDLVLEGDS CVTIMSKDKPTIDVKMMNMEAA
NLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWNGCGLFG
KGSIDTCAKFACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYPTQVGATQAGRF
SITPAAPSYTLKLGEYGEVTVDCPRSGIDTNAYYVMTVGTKTFLVHREWFMDLNL PW
SSAGSTVWRNRETLM EFEEPHATKQSVIALGSQEGALHQALAGAI PVEFSSNTVKLTS
GHLKCRVKMEKLQLKGT TYGVCSKAFKFLGTPADTGHGT VVLELQYTGT DGPCKVPIS
SVASLNDLTPVGR LVTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQINHHWHK
SGSSIGKAFTTTLKGAQRLAALGDTAWDFGSGGVFTSVGKAVHQVFEGGAFRSLFGGM
SWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHA"
mat_peptide 1. .276
/product="pre-membrane protein prM"
mat_peptide 277. .501
/product="membrane protein M"
mat_peptide 502. .2004
/product="envelope protein E"

Query Match		65.6%;	Score 997.2;	DB 10;	Length 2004;
Best Local Similarity		78.7%;	Pred. No. 0;		
Matches 1191;		Conservative	0;	Mismatches 323;	Indels 0; Gaps 0;
Qy	1	CGGAATTCAGCTTCAACTGTTTAGGAATGAGCAACAGGGACTTCCTGGAGGGAGTGCTCTG	60		
Db	491	CAGCTTACAGCTTCAACTGCCTTGAATGAGCAACAGAGACTTCTTGGAAGGAGTGCTCTG	550		
Qy	61	GAGCTACATGGGTGATCTGGTACTGGAAGGAGACAGTTGTGTGACCATAATGTCAAAG	120		
Db	551	GAGCAACATGGGTGGATTGTTCTCGAAGGCGACAGCTGCGTGACTATCATGTCTAAGG	610		
Qy	121	ACAAGCCAACCATTGATGTCAAAATGATGAACATGGAAGCAGCTAATCTCGCAGATGTGC	180		
Db	611	ACAAGCCTACCATCGATGTGAAGATGATGAATATGGAGGCGGCCAACCTGGCAGAGGTCC	670		
Qy	181	GTAGCTACTGCTACTTAGCTTCGGTCAGTGATCTGTCAACAAAAGCCGCGTGTCCAACCA	240		
Db	671	GTAGTTATTGCTATTTGGCTACCGTCAGCGATCTCTCCACCAAAGCTGCGTGCCCCACCA	730		
Qy	241	TGGGTGAAGCTCACAAAGAGAAAAGAGCCGACCCTGCCTTTGTTTTGCAAGCAAGGCGTCG	300		
Db	731	TGGGAGAAGCTCACAAATGACAAACGTGCTGACCCAGCTTTTGTGTGCAGACAAGGAGTGG	790		
Qy	301	TAGACAGAGGATGGGGGAATGGATGCGGACTGTTTGGAAAGGGGAGCATTGACACATGTG	360		
Db	791	TGGACAGGGGCTGGGGCAACGGCTGCGGACTATTTGGCAAAGGAAGCATTGACACATGCG	850		
Qy	361	CAAAGTTTGCCTGTACAACCAAGGCAACTGGTTGGATTATCCAGAAGGAAAACATCAAGT	420		
Db	851	CCAAATTTGCCTGCTCTACCAAGGCAATAGGAAGAACCATCTTGAAAGAGAATATCAAGT	910		
Qy	421	ACGAGGTTGCCATATTTGTGCATGGCCCGACGACTGTGGAATCACATGGCAATTATTCAA	480		
Db	911	ACGAAGTGGCCATTTTTGTCCATGGACCAACTACTGTGGAGTCGCACGGAACTACCCCA	970		
Qy	481	CACAGATAGGGGCTACCCAAGCAGGAAGGTTGAGCATAACTCCATCGGCACCATCCTACA	540		
Db	971	CACAGGTTGGAGCCACTCAGGCAGGGAGATTGAGCATCACTCCTGCGGCGCCTTCATACA	1030		
Qy	541	CGCTGAAGTTGGGTGAGTATGGTGAGGTCACAGTTGACTGTGAGCCACGGTCAGGAATAG	600		
Db	1031	CACTAAAGCTTGAGAATATGGAGAGGTGACAGTGGACTGTGAACCACGGTCAGGGATTG	1090		
Qy	601	ACACTAGCGCTTACTACGTTATGTCAGTGGGTGCGAAGTCCTTCTTGTTTACCGAGAAT	660		
Db	1091	ACACCAATGCATACTACGTGATGACTGTTGGAACAAAGACGTTCTTGTTCCATCGTGAGT	1150		
Qy	661	GGTTTATGGACCTGAACCTTCCATGGAGTAGCGCTGGAAGCACAACTGGAGGAACCGGG	720		
Db	1151	GGTTCATGGATCTCAACCTCCCTTGAGCAGTGCTGGAAGTACTGTGTGGAGGAACAGAG	1210		
Qy	721	AAACACTGATGGAGTTTGAAGAACCCTCATGCCACCAAACAATCTGTCGTAGCTCTAGGGT	780		
Db	1211	AGACGTTAATGGAGTTTGAAGAACACACGCCACGAAGCAGTCTGTGATAGCATTGGGCT	1270		

DEFINITION West Nile virus polyprotein gene, partial cds.
 ACCESSION AY963774
 VERSION AY963774.1 GI:63098701
 KEYWORDS .
 SOURCE West Nile virus (WNV)
 ORGANISM West Nile virus
 Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 Flavivirus; Japanese encephalitis virus group.
 REFERENCE 1 (bases 1 to 2004)
 AUTHORS Elizondo-Quiroga,D., Davis,C.T., Fernandez-Salas,I.,
 Escobar-Lopez,R., Velazco Olmos,D., Soto Gastelum,L.C., Aviles
 Acosta,M., Elizondo-Quiroga,A., Gonzalez-Rojas,J.I., Contreras
 Cordero,J.F., Guzman,H., Travassos da Rosa,A., Blitvich,B.J.,
 Barret,A.D., Beaty,B.J. and Tesh,R.B.
 TITLE West Nile Virus Isolation in Human and Mosquitoes, Mexico
 JOURNAL Emerging Infect. Dis. 11 (9), 1449-1452 (2005)
 REFERENCE 2 (bases 1 to 2004)
 AUTHORS Elizondo-Quiroga,D., Davis,C., Fernandez-Salas,I.,
 Escobar-Lopez,R., Velazco-Olmos,D., Soto-Gastelum,L.,
 Aviles-Acosta,M., Elizondo-Quiroga,A., Gonzalez-Rojas,J.,
 Contreras-Cordero,J., Guzman,H., Travassos da Rosa,A., Blitvich,B.,
 Beaty,B., Barret,A. and Tesh,R.
 TITLE Direct Submission
 JOURNAL Submitted (16-MAR-2005) Pathology, University of Texas Medical
 Branch, 301 University Ave., Galveston, TX 77555, USA
 FEATURES
 source Location/Qualifiers
 1. .2004
 /organism="West Nile virus"
 /mol_type="genomic RNA"
 /serotype="lineage I"
 /isolation_source="serum"
 /specific_host="Homo sapiens"
 /db_xref="taxon:11082"
 /country="Mexico: Sonora"
 CDS
 <1..>2004
 /codon_start=1
 /product="polyprotein"
 /protein_id="AAY32589.1"
 /db_xref="GI:63098702"
 /translation="VTLSNFQGKVMNTVNATDVTDVITIPTAAGKNLCIVRAMDVGYM
 CDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRSLTVQTHGES
 TLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRVVFWLLL
 LVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVKMMNMEAA
 NLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWGNGCGLFG
 KGSIDTCAKFACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYSTQAGATQAGRF
 SITPAAPSYTLKLGEYGEVTVDCPRSGIDTNAYYVMTVGTKTFLVHREWFDLNLPLW
 SSAGSTVWRNRETILMEFEEPHATKQSVIALGSEQEGALHQALAGAIPEVFSNTVKLTS
 GHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTGDPCKVPIS
 SVASLNDLTPVGRLLVTNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQINHHWHK
 SGSSIGKAFTTTLKGAQRLAALGDTAWDFGSGVGVFTSVGKAVHQVFGGAFRSLFGGM
 SWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHA"
 mat_peptide 466..966
 /product="pre-membrane protein; prM/M"
 mat_peptide 967..>2004
 /product="envelope protein"
 ORIGIN

Query Match 65.6%; Score 997.2; DB 10; Length 2004;
Best Local Similarity 78.7%; Pred. No. 0;
Matches 1191; Conservative 0; Mismatches 323; Indels 0; Gaps 0;

```
Qy      1 CGGAATTCAGCTTCAACTGTTTAGGAATGAGCAACAGGGACTTCCTGGAGGGAGTGTCTG 60
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     491 CAGCTTACAGCTTCAACTGCCCTTGAATGAGCAACAGAGACTTCTTGAAGGAGTGTCTG 550

Qy      61 GAGCTACATGGGTTGATCTGGTACTGGAAGGAGACAGTTGTGTGACCATAATGTCAAAAG 120
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     551 GAGCAACATGGGTGGATTGTTCTCGAAGGCGACAGCTGCGTGACTATCATGTCTAAAG 610

Qy     121 ACAAGCCAACCATTGATGTCAAAATGATGAACATGGAAGCAGCTAATCTCGCAGATGTGC 180
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     611 ACAAGCCTACCATCGATGTGAAGATGATGAATATGGAGGCGGCCAACCTGGCAGAGGTCC 670

Qy     181 GTAGCTACTGCTACTTAGCTTCGGTCAGTGATCTGTCAACAAAAGCCGCGTGCCAACCA 240
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     671 GCAGTTATTGCTATTTGGCTACCGTCAGCGATCTCTCCACCAAAGCTGCGTGCCCGACCA 730

Qy     241 TGGGTGAAGCTCACAACGAGAAAAGAGCCGACCCTGCCTTTGTTTGCAAGCAAGGCGTCG 300
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     731 TGGGAGAAGCTCACAATGACAAACGTGCTGACCCAGCTTTTGTGTGCAGACAAGGAGTGG 790

Qy     301 TAGACAGAGGATGGGGGAATGGATGCGGACTGTTTGGAAGGGGAGCATTGACACATGTG 360
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     791 TGGACAGGGGCTGGGGCAACGGCTGCGGACTATTTGGCAAAGGAAGCATTGACACATGCG 850

Qy     361 CAAAGTTTGCCTGTACAACCAAGGCAACTGGTTGGATTATCCAGAAGGAAAACATCAAGT 420
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     851 CCAAGTTTGCCTGCTCTACCAAGGCAATAGGAAGAACCATCTTGAAAGAGAATATCAAGT 910

Qy     421 ACGAGGTTGCCATATTTGTGCATGGCCCGACGACTGTGGAATCACATGGCAATTATTCAA 480
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     911 ACGAAGTGGCCATTTTGTCCATGGACCAACTACTGTGGAGTCGCACGGAACTACTCCA 970

Qy     481 CACAGATAGGGGCTACCCAAGCAGGAAGGTTTACGCATAACTCCATCGGCACCATCCTACA 540
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     971 CACAGGCTGGAGCCACTCAGGCAGGGAGATTACAGCATCACTCCTGCGGCGCCTTCATACA 1030

Qy     541 CGCTGAAGTTGGGTGAGTATGGTGAGGTCACAGTTGACTGTGAGCCACGGTCAGGAATAG 600
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    1031 CACTAAAGCTTGAGAAATATGGAGAGGTGACAGTGGACTGTGAACCACGGTCAGGGATTG 1090

Qy     601 AACTAGCGCTTACTACGTTATGTCACTGGGTGCGAAGTCCTTCTTGGTTCACCGAGAAT 660
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    1091 ACACCAATGCATACTACGTGATGACTGTTGGAACAAAGACGTTCTTGGTCCATCGTGAGT 1150

Qy     661 GGTTTATGGACCTGAACCTTCCATGGAGTAGCGCTGGAAGCACAACGTGGAGGAACCGGG 720
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    1151 GGTTTATGGACCTCAACCTCCCTTGGAGCAGTGCTGGAAGCACTGTGTGGAGGAACAGAG 1210

Qy     721 AAACACTGATGGAGTTTGAAGAACCTCATGCCACCAAACAATCTGTCTAGCTCTAGGGT 780
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    1211 AGACGTTAATGGAGTTTGAAGAACCACACGCCACGAAGCAGTCTGTGATAGCATTGGGCT 1270

Qy     781 CGCAGGAAGGTGCCTTGCACCAAGCTCTGGCTGGAGCAATTCCTGTTGAGTTCTCAAGCA 840
```


DR GO; GO:0019031; C:viral envelope; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
 DR GO; GO:0003725; F:double-stranded RNA binding; IEA.
 DR GO; GO:0003724; F:RNA helicase activity; IEA.
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
 DR GO; GO:0004252; F:serine-type endopeptidase activity; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0019079; P:viral genome replication; IEA.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR011545; DEAD/DEAH_N.
 DR InterPro; IPR011999; Flav_glyE_cen_dm.
 DR InterPro; IPR001122; Flavi_capsidC.
 DR InterPro; IPR011492; Flavi_DEAD.
 DR InterPro; IPR000069; Flavi_M.
 DR InterPro; IPR001157; Flavi_NS1.
 DR InterPro; IPR000752; Flavi_NS2A.
 DR InterPro; IPR000487; Flavi_NS2B.
 DR InterPro; IPR000404; Flavi_NS4A.
 DR InterPro; IPR001528; Flavi_NS4B.
 DR InterPro; IPR000208; Flavi_NS5.
 DR InterPro; IPR002535; Flavi_propep.
 DR InterPro; IPR000336; Flv_glyE_Ig-like.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR001850; Peptidase_S7.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSwir.
 DR InterPro; IPR002877; RrmJFtsJ_mtfrase.
 DR InterPro; IPR011998; Vrl_glyE_cen_dim.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF01003; Flavi_capsid; 1.
 DR Pfam; PF07652; Flavi_DEAD; 1.
 DR Pfam; PF02832; Flavi_glycop_C; 1.
 DR Pfam; PF00869; Flavi_glycoprot; 1.
 DR Pfam; PF01004; Flavi_M; 1.
 DR Pfam; PF00948; Flavi_NS1; 1.
 DR Pfam; PF01005; Flavi_NS2A; 1.
 DR Pfam; PF01002; Flavi_NS2B; 1.
 DR Pfam; PF01350; Flavi_NS4A; 1.
 DR Pfam; PF01349; Flavi_NS4B; 1.
 DR Pfam; PF00972; Flavi_NS5; 1.
 DR Pfam; PF01570; Flavi_propep; 1.
 DR Pfam; PF01728; FtsJ; 1.
 DR Pfam; PF00271; Helicase_C; 1.
 DR Pfam; PF00949; Peptidase_S7; 1.
 DR ProDom; PD001496; Flavi_NS1; 1.
 DR SMART; SM00487; DEXDc; 1.
 DR SMART; SM00490; HELICc; 1.
 DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.
 KW Polyprotein.
 FT CHAIN 1 123 capsid.
 FT CHAIN 124 215 prM.
 FT CHAIN 216 290 M.
 FT CHAIN 291 791 envelope.
 FT CHAIN 792 1144 NS1.
 FT CHAIN 1145 1374 NS2A.
 FT CHAIN 1375 1505 NS2B.

FT CHAIN 1506 2124 NS3.
 FT CHAIN 2125 2273 NS4A.
 FT CHAIN 2274 2528 NS4B.
 FT CHAIN 2529 3433 NS5.
 SQ SEQUENCE 3433 AA; 381124 MW; C302F24541A66BC8 CRC64;

Query Match 95.9%; Score 2531; DB 2; Length 3433;
 Best Local Similarity 95.4%; Pred. No. 3.8e-183;
 Matches 478; Conservative 14; Mismatches 9; Indels 0; Gaps 0;

```

Qy      1 FNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVKMMNMEAANLADVRSYC 60
      |||:|||||
Db     291 FNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVKMMNMEAANLAEVRSYC 350

Qy     61 YLASVSDLSTKAACPTMGEAHNEKRADPAFVCKQGVVDRGWNGCGLFGKGSIDTCAKFA 120
      |||:|||||
Db    351 YLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWNGCGLFGKGSIDTCAKFA 410

Qy    121 CTTKATGWIIQKENIKYEVAIFVHGPTTVESHGNYSTQIGATQAGRFSITPSAPSYTLKL 180
      |:|||||
Db    411 CSTKATGRTILKENIKYEVAIFVHGPTTVESHGNYSTQIGATQAGRFSITPAAPSYTLKL 470

Qy    181 GEYGEVTVDCEPRSGIDTSAYYVMSVGAKSFLVHREWFMDLNLPWSSAGSTTWNRNRETLM 240
      |||:|||||
Db    471 GEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWFMDLNLPWSSAGSTVWRNRETLM 530

Qy    241 EFEEPHATKQSVVALGSQEGALHQALAGAI PVEFSSNTVKLTSGHLKCRVKMEKLQLKGT 300
      |||:|||||
Db    531 EFEEPHATKQSVIALGSQEGALHQALAGAI PVEFSSNTVKLTSGHLKCRVKMEKLQLKGT 590

Qy    301 TYGVCSKAFKFARTPADTGHGTVVLELQYTGKDGPCCKVPISSVASLNDLTPVGRLVTVNP 360
      |||:|||||
Db    591 TYGVCSKAFKFLGTPADTGHGTVVLELQYTGTGDPCKVPISSVASLNDLTPVGRLVTVNP 650

Qy    361 FVSVATANSKVLIELEPPFSDSYIVVGRGEQQINHHWHKSGSSIGKAFTTTTLRGAQRLAA 420
      |||:|||||
Db    651 FVSVATANAKVLIELEPPFGDSYIVVGRGEQQINHHWHKSGSSIGKAFTTTTLKGAQRLAA 710

Qy    421 LGDTAWDFGSGGVFTSVGKAIHQVFGGAFRSLFGGMSWITQGLLGALLLWMGINARDS 480
      |||:|||||
Db    711 LGDTAWDFGSGGVFTSVGKAVHQVFGGAFRSLFGGMSWITQGLLGALLLWMGINARDS 770

Qy    481 IAMTFLAVGGVLLFLSVNVHA 501
      ||:|||||
Db    771 IALTFLAVGGVLLFLSVNVHA 791

```

RESULT 13

Q8JU42_WNV

ID Q8JU42_WNV PRELIMINARY; PRT; 3433 AA.

AC Q8JU42;

DT 01-OCT-2002, integrated into UniProtKB/TrEMBL.

DT 01-OCT-2002, sequence version 1.

DT 07-FEB-2006, entry version 14.

DE Polyprotein.

OS West Nile virus (WN).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Flavivirus; Japanese encephalitis virus group.
 OX NCBI_TaxID=11082;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RX MEDLINE=22089180; PubMed=12093177; DOI=10.1006/viro.2002.1449;
 RA Lanciotti R.S., Ebel G.D., Deubel V., Kerst A.J., Murri S., Meyer R.,
 RA Bowen M., McKinney N., Morrill W.E., Crabtree M.B., Kramer L.D.,
 RA Roehrig J.T.;
 RT "Complete genome sequences and phylogenetic analysis of West Nile
 RT virus strains isolated from the United States, Europe, and the Middle
 RT East.";
 RL Virology 298:96-105(2002).
 CC -----
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License
 CC -----
 DR EMBL; AF404757; AAM81753.1; -; Genomic_RNA.
 DR HSSP; Q88653; 1L9K.
 DR SMR; Q8JU42; 25-97.
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0019031; C:viral envelope; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
 DR GO; GO:0003725; F:double-stranded RNA binding; IEA.
 DR GO; GO:0003724; F:RNA helicase activity; IEA.
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
 DR GO; GO:0004252; F:serine-type endopeptidase activity; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0019079; P:viral genome replication; IEA.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR011545; DEAD/DEAH_N.
 DR InterPro; IPR011999; Flav_glyE_cen_dm.
 DR InterPro; IPR001122; Flavi_capsidC.
 DR InterPro; IPR011492; Flavi_DEAD.
 DR InterPro; IPR000069; Flavi_M.
 DR InterPro; IPR001157; Flavi_NS1.
 DR InterPro; IPR000752; Flavi_NS2A.
 DR InterPro; IPR000487; Flavi_NS2B.
 DR InterPro; IPR000404; Flavi_NS4A.
 DR InterPro; IPR001528; Flavi_NS4B.
 DR InterPro; IPR000208; Flavi_NS5.
 DR InterPro; IPR002535; Flavi_propep.
 DR InterPro; IPR000336; Flv_glyE_Ig-like.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR001850; Peptidase_S7.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSvir.
 DR InterPro; IPR002877; RrmJFtsJ_mtfrase.
 DR InterPro; IPR011998; Vrl_glyE_cen_dim.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF01003; Flavi_capsid; 1.
 DR Pfam; PF07652; Flavi_DEAD; 1.
 DR Pfam; PF02832; Flavi_glycop_C; 1.
 DR Pfam; PF00869; Flavi_glycoprot; 1.
 DR Pfam; PF01004; Flavi_M; 1.
 DR Pfam; PF00948; Flavi_NS1; 1.

DR Pfam; PF01005; Flavi_NS2A; 1.
 DR Pfam; PF01002; Flavi_NS2B; 1.
 DR Pfam; PF01350; Flavi_NS4A; 1.
 DR Pfam; PF01349; Flavi_NS4B; 1.
 DR Pfam; PF00972; Flavi_NS5; 1.
 DR Pfam; PF01570; Flavi_propep; 1.
 DR Pfam; PF01728; FtsJ; 1.
 DR Pfam; PF00271; Helicase_C; 1.
 DR Pfam; PF00949; Peptidase_S7; 1.
 DR ProDom; PD001496; Flavi_NS1; 1.
 DR SMART; SM00487; DEXDc; 1.
 DR SMART; SM00490; HELICc; 1.
 DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.
 KW Polyprotein.
 SQ SEQUENCE 3433 AA; 381210 MW; 1DFFCCDB2174B7EE CRC64;

Query Match 95.9%; Score 2531; DB 2; Length 3433;
 Best Local Similarity 95.4%; Pred. No. 3.8e-183;
 Matches 478; Conservative 14; Mismatches 9; Indels 0; Gaps 0;

Qy	1	FNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVKMMNMEAANLADVRSYC	60
		:	
Db	291	FNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVKMMNMEAANLAEVRSYC	350
Qy	61	YLASVSDLSTKAACPTMGEAHNEKRA DPAFVCKQG VVDRGWNGCGLFGKGSIDTCAKFA	120
		: : :	
Db	351	YLATVSDLSTKAACPTMGEAHNDKRA DPAFVCRQG VVDRGWNGCGLFGKGSIDTCAKFA	410
Qy	121	CTTKATGWIIQKENIKYEVAIFVHGPTTVESHGNYSTQIGATQAGRSITPSAPSYTLKL	180
		: : :	
Db	411	CSTKATGRTILKENIKYEVAIFVHGPTTVESHGNYSTQIGATQAGRSITPAAPSYTLKL	470
Qy	181	GEYGEVTV DCEPRSGIDTSAYYVMSVGAKSFLVHREWFMDLNL PWSSAGSTTWRNRETLM	240
		: : :	
Db	471	GEYGEVTV DCEPRSGIDTNAYYVMTVGTKTFLVHREWFMDLNL PWSSAGSTVWRNRETLM	530
Qy	241	EFEEPHATKQSVVALGSQEGALHQALAGAI PVEFSSNTVKLTSGHLKCRVKMEKLQLKGT	300
		: :	
Db	531	EFEEPHATKQSVIALGSQEGALHQALAGAI PVEFSSNTVKLTSGHLKCRVKMEKLQLKGT	590
Qy	301	TYGVC SKAFKFARTPADTGHGTVVLELQYTGKDG PCKVPISSVASLNDLTPVGRLVTNP	360
		:	
Db	591	TYGVC SKAFKFLGTPADTGHGTVVLELQYTGT DGPCKVPISSVASLNDLTPVGRLVTNP	650
Qy	361	FVSVATANSKVLIELEPPFSDSYIVVGRGEQQINHHWHKSGSSIGKAFTTTLRGAQRLAA	420
		: :	
Db	651	FVSVATANAKVLIELEPPFGDSYIVVGRGEQQINHHWHKSGSSIGKAFTTTLKGAQRLAA	710
Qy	421	LGDTAWDFG SVGGVFTSVGKAIHQVF GGAFRSLFGGMSWITQGLLGALLLWMGINARDS	480
		:	
Db	711	LGDTAWDFG SVGGVFTSVGKAVHQVF GGAFRSLFGGMSWITQGLLGALLLWMGINARDS	770
Qy	481	IAMTFLAVGGVLLFLSVNVHA	501
		:	
Db	771	IALTFLAVGGVLLFLSVNVHA	791

RESULT 14

Q9EA21_WNV

ID Q9EA21_WNV PRELIMINARY; PRT; 3433 AA.

AC Q9EA21;

DT 01-MAR-2001, integrated into UniProtKB/TrEMBL.

DT 01-MAR-2001, sequence version 1.

DT 07-FEB-2006, entry version 20.

DE Polyprotein.

OS West Nile virus (WN).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Flavivirus; Japanese encephalitis virus group.

OX NCBI_TaxID=11082;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=RO97-50;

RX MEDLINE=20014331; PubMed=10548295;

RA Savage H.M., Ceianu C., Nicolescu G., Karabatsos N., Lanciotti R.,

RA Vladimirescu A., Laiv L., Ungureanu A., Romanca C., Tsai T.F.;

RT "Entomologic and avian investigations of an epidemic of West Nile

RT fever in Romania in 1996, with serologic and molecular

RT characterization of a virus isolate from mosquitoes.";

RL Am. J. Trop. Med. Hyg. 61:600-611(1999).

RN [2]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=RO97-50;

RA Bowen M., Meyer R.F., McKinney N., Morrill W., Lanciotti R.;

RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.

CC

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>

CC Distributed under the Creative Commons Attribution-NoDerivs License

CC

DR EMBL; AF260969; AAG02040.1; -; Genomic_RNA.

DR HSSP; Q88653; 1L9K.

DR SMR; Q9EA21; 25-97.

DR GO; GO:0016021; C:integral to membrane; IEA.

DR GO; GO:0019028; C:viral capsid; IEA.

DR GO; GO:0019031; C:viral envelope; IEA.

DR GO; GO:0005524; F:ATP binding; IEA.

DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.

DR GO; GO:0003725; F:double-stranded RNA binding; IEA.

DR GO; GO:0003724; F:RNA helicase activity; IEA.

DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.

DR GO; GO:0004252; F:serine-type endopeptidase activity; IEA.

DR GO; GO:0005198; F:structural molecule activity; IEA.

DR GO; GO:0019079; P:viral genome replication; IEA.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR011545; DEAD/DEAH_N.

DR InterPro; IPR011999; Flav_glyE_cen_dm.

DR InterPro; IPR001122; Flavi_capsidC.

DR InterPro; IPR011492; Flavi_DEAD.

DR InterPro; IPR000069; Flavi_M.

DR InterPro; IPR001157; Flavi_NS1.

DR InterPro; IPR000752; Flavi_NS2A.

DR InterPro; IPR000487; Flavi_NS2B.

DR InterPro; IPR000404; Flavi_NS4A.

DR InterPro; IPR001528; Flavi_NS4B.

DR InterPro; IPR000208; Flavi_NS5.


```

      ||||||||||||||||:|||||||||||||||||||||||
Db      686 LGDTAWDFGSVGGVFTSVGKAVHQVFGGAFRSLFGGMSWITQGLLGALLLWMGINARDS 745

Qy      481 IAMTFLAVGGVLLFLSVNVHA 501
      ||:|||||||||||||||
Db      746 IALTFLAVGGVLLFLSVNVHA 766

```

Search completed: June 10, 2006, 02:43:07
Job time : 303 secs

Db 651 FVSVATANSKVLIELEPPFGDSYIVVGRGEQQINHHWHKSGSSIGKAFTTTTLRGAQRLAA 710

Qy 421 LGDTAWDFGSVGGVFTSVGKAIHQVFGGAFRSLFGGMSWITQGLLGALLLWMGINARDS 480
 |||

Db 711 LGDTAWDFGSVGGVFTSVGKAIHQVFGGAFRSLFGGMSWITQGLLGALLLWMGINARDS 770

Qy 481 IAMTFLAVGGVLLFLSVNVHA 501
 |||

Db 771 IAMTFLAVGGVLLFLSVNVHA 791

RESULT 4

Q5MXE3_WNV

ID Q5MXE3_WNV PRELIMINARY; PRT; 3430 AA.

AC Q5MXE3;

DT 01-FEB-2005, integrated into UniProtKB/TrEMBL.

DT 01-FEB-2005, sequence version 1.

DT 07-FEB-2006, entry version 4.

DE Polyprotein.

OS West Nile virus (WN).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Flavivirus; Japanese encephalitis virus group.

OX NCBI_TaxID=11082;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=B956;

RX PubMed=15527855; DOI=10.1016/j.virol.2004.09.014;

RA Yamshchikov G., Borisevich V., Seregin A., Chaporgina E., Mishina M.,

RA Mishin V., Wai Kwok C., Yamshchikov V.;

RT "An attenuated West Nile prototype virus is highly immunogenic and

RT protects against the deadly NY99 strain: a candidate for live WN

RT vaccine development.";

RL Virology 330:304-312(2004).

RN [2]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=B956;

RA Borisevich V.G., Yamshchikov V.F.;

RT "Molecular basis of attenuation of the West Nile virus prototype

RT strain B956.";

RL Submitted (JAN-2004) to the EMBL/GenBank/DDBJ databases.

CC -----

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>

CC Distributed under the Creative Commons Attribution-NoDerivs License

CC -----

DR EMBL; AY532665; AAT02759.1; -; Genomic_RNA.

DR SMR; Q5MXE3; 25-97.

DR GO; GO:0016021; C:integral to membrane; IEA.

DR GO; GO:0019028; C:viral capsid; IEA.

DR GO; GO:0019031; C:viral envelope; IEA.

DR GO; GO:0005524; F:ATP binding; IEA.

DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.

DR GO; GO:0003725; F:double-stranded RNA binding; IEA.

DR GO; GO:0003724; F:RNA helicase activity; IEA.

DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.

DR GO; GO:0004252; F:serine-type endopeptidase activity; IEA.

DR GO; GO:0005198; F:structural molecule activity; IEA.

DR GO; GO:0019079; P:viral genome replication; IEA.

DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR011545; DEAD/DEAH_N.
 DR InterPro; IPR011999; Flav_glyE_cen_dm.
 DR InterPro; IPR001122; Flavi_capsidC.
 DR InterPro; IPR011492; Flavi_DEAD.
 DR InterPro; IPR000069; Flavi_M.
 DR InterPro; IPR001157; Flavi_NS1.
 DR InterPro; IPR000752; Flavi_NS2A.
 DR InterPro; IPR000487; Flavi_NS2B.
 DR InterPro; IPR000404; Flavi_NS4A.
 DR InterPro; IPR001528; Flavi_NS4B.
 DR InterPro; IPR000208; Flavi_NS5.
 DR InterPro; IPR002535; Flavi_propep.
 DR InterPro; IPR000336; Flv_glyE_Ig-like.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR001850; Peptidase_S7.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR InterPro; IPR002877; RrmJFtsJ_mtfrase.
 DR InterPro; IPR011998; Vrl_glyE_cen_dim.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF01003; Flavi_capsid; 1.
 DR Pfam; PF07652; Flavi_DEAD; 1.
 DR Pfam; PF02832; Flavi_glycop_C; 1.
 DR Pfam; PF00869; Flavi_glycoprot; 1.
 DR Pfam; PF01004; Flavi_M; 1.
 DR Pfam; PF00948; Flavi_NS1; 1.
 DR Pfam; PF01005; Flavi_NS2A; 1.
 DR Pfam; PF01002; Flavi_NS2B; 1.
 DR Pfam; PF01350; Flavi_NS4A; 1.
 DR Pfam; PF01349; Flavi_NS4B; 1.
 DR Pfam; PF00972; Flavi_NS5; 1.
 DR Pfam; PF01570; Flavi_propep; 1.
 DR Pfam; PF01728; FtsJ; 1.
 DR Pfam; PF00271; Helicase_C; 1.
 DR Pfam; PF00949; Peptidase_S7; 1.
 DR ProDom; PD001496; Flavi_NS1; 1.
 DR SMART; SM00487; DEXDc; 1.
 DR SMART; SM00490; HELICc; 1.
 DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.
 KW Polyprotein.
 SQ SEQUENCE 3430 AA; 379894 MW; 6298C302480200D8 CRC64;

Query Match 97.6%; Score 2575; DB 2; Length 3430;
 Best Local Similarity 98.2%; Pred. No. 1.7e-186;
 Matches 492; Conservative 3; Mismatches 2; Indels 4; Gaps 1;

Qy 1 FNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVKMMNMEAANLADVRSYC 60
 |||:|||||
 Db 291 FNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTILMSKDKPTIDVKMMNMEAANLADVRSYC 350
 Qy 61 YLASVSDLSTKAACPTMGEAHNEKRADPAFVCKQGVVDRGWNGCGLFGKGSIDTCAKFA 120
 |||:|||||
 Db 351 YLASVSDLSTRAACPTMGEAHNEKRADPAFVCKQGVVDRGWNGCGLFGKGSIDTCAKFA 410
 Qy 121 CTTKATGWIIQKENIKYEVAIFVHGPTTVESHGNYSTQIGATQAGRFSITPSAPSYTLKL 180
 |||:|||||

Db 411 CTTKATGWIIQKENIKYEVAIFVHGPTTVESHG----KIGATQAGRFSITPSAPSYTLKL 466

Qy 181 GEYGEVTVDCEPRSGIDTSAYYVMSVGAKSFLVHREWFMDLNLPWSSAGSTTWRNRETLM 240
 |||

Db 467 GEYGEVTVDCEPRSGIDTSAYYVMSVGAKSFLVHREWFMDLNLPWSSAGSTTWRNRETLM 526

Qy 241 EFEEPHATKQSVVALGSQEGALHQALAGAIPEFSSNTVKLTSGHLKCRVKMEKLQLKGT 300
 |||

Db 527 EFEEPHATKQSVVALGSQEGALHQALAGAIPEFSSNTVKLTSGHLKCRVKMEKLQLKGT 586

Qy 301 TYGVCSKAFKFARTPADTGHGTVVLELQYTGKDGPKVPISSVASLNDLTPVGRLVTVNP 360
 |||

Db 587 TYGVCSKAFKFARTPADTGHGTVVLELQYTGKDGPKVPISSVASLNDLTPVGRLVTVNP 646

Qy 361 FVSVATANSKVLIELEPPFSDSYIVVGRGEQQINHHWHKSGSSIGKAFTTTTLRGAQRLAA 420
 |||

Db 647 FVSVATANSKVLIELEPPFSDSYIVVGRGEQQINHHWHKSGSSIGKAFTTTTLRGAQRLAA 706

Qy 421 LGDTAWDFGSGGVFTSVGKAIHQVFGGAFRSLFGGMSWITQGLLGALLLWMGINARDS 480
 |||

Db 707 LGDTAWDFGSGGVFTSVGKAIHQVFGGAFRSLFGGMSWITQGLLGALLLWMGINARDS 766

Qy 481 IAMTFLAVGGVLLFLSVNVHA 501
 |||

Db 767 IAMTFLAVGGVLLFLSVNVHA 787

RESULT 5

Q2PMF5_WNV

ID Q2PMF5_WNV PRELIMINARY; PRT; 3430 AA.

AC Q2PMF5;

DT 24-JAN-2006, integrated into UniProtKB/TrEMBL.

DT 24-JAN-2006, sequence version 1.

DT 07-FEB-2006, entry version 2.

DE Polyprotein precursor.

OS West Nile virus (WN).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Flavivirus; Japanese encephalitis virus group.

OX NCBI_TaxID=11082;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=ArD76104;

RA Borisevich V.G., Seregin A.V., Yamshchikov V.F.;

RT "Genetic determinants of West Nile virus pathogenicity.";

RL Submitted (DEC-2005) to the EMBL/GenBank/DDBJ databases.

CC -----

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>

CC Distributed under the Creative Commons Attribution-NoDerivs License

CC -----

DR EMBL; DQ318019; ABC49716.1; -; mRNA.

KW Polyprotein; Signal.

FT SIGNAL 106 123 Potential.

FT SIGNAL 275 290 Potential.

FT SIGNAL 764 787 Potential.

FT SIGNAL 2247 2269 Potential.

FT CHAIN 1 105 C protein.

FT CHAIN 124 215 cleaved amino terminal prM fragment.

FT	CHAIN	124	290	prM protein.
FT	CHAIN	216	290	M protein.
FT	CHAIN	291	787	E protein.
FT	CHAIN	788	1139	NS1 protein.
FT	CHAIN	1140	1370	NS2A protein.
FT	CHAIN	1371	1501	NS2B protein.
FT	CHAIN	1502	2120	NS3 protein.
FT	CHAIN	2121	2269	NS4A protein.
FT	CHAIN	2270	2525	NS4B protein.
FT	CHAIN	2526	3430	NS5 protein.
SQ	SEQUENCE	3430 AA;	379866 MW;	B03CBB31C86FD33B CRC64;

Query Match 97.5%; Score 2573; DB 2; Length 3430;
 Best Local Similarity 98.2%; Pred. No. 2.4e-186;
 Matches 492; Conservative 3; Mismatches 2; Indels 4; Gaps 1;

Qy	1	FNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVKMMNMEAANLADVRSYC	60
Db	291	FNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVKMMNMEAANLADVRSYC	350
Qy	61	YLASVSDLSTKAACPTMGEAHNEKRADPAFVCKQGVVDRGWNGCGLFGKGSIDTCAKFA	120
Db	351	YLASVSDLSTRAACPTMGEAHNEKRADPAFVCKQGVVDRGWNGCGLFGKGSIDTCAKFA	410
Qy	121	CTTKATGWIIQKENIKYEVAIFVHGPTTVESHGNYSTQIGATQAGRFSITPSAPSYTLKL	180
Db	411	CTTKATGWIIQKENIKYEVAIFVHGPTTVESHG----KIGATQAGRFSITPSAPSYTLKL	466
Qy	181	GEYGEVTVDCPRSGIDTSAYYVMSVGAKSFLVHREWFMDLNLPWSSAGSTTWRNRETLN	240
Db	467	GEYGEVTVDCPRSGIDTSAYYVMSVGAKSFLVHREWFMDLNLPWSSAGSTTWRNRETLV	526
Qy	241	EFEEPHATKQSVVALGSQEGALHQALAGAIPEFSSNTVKLTSGHLKCRVKMEKLQLKGT	300
Db	527	EFEEPHATKQSVVALGSQEGALHQALAGAIPEFSSNTVKLTSGHLKCRVKMEKLQLKGT	586
Qy	301	TYGVCSKAFKFARTPADTGHGTVVLELQYTGKDGPKVPISSVASLNDLTPVGRLVTVNP	360
Db	587	TYGVCSKAFKFARTPADTGHGTVVLELQYTGTDGPKVPISSVASLNDLTPVGRLVTVNP	646
Qy	361	FVSVATANSKVLIELEPPFSDSYIVVGRGEQQINHHWHKSGSSIGKAFTTTTLRGAQRLAA	420
Db	647	FVSVATANSKVLIELEPPFGDSYIVVGRGEQQINHHWHKSGSSIGKAFTTTTLRGAQRLAA	706
Qy	421	LGDTAWDFGSVGGVFTSVGKAIHQVFGGAFRSLFGGMSWITQGLLGALLLWMGINARDS	480
Db	707	LGDTAWDFGSVGGVFTSVGKAIHQVFGGAFRSLFGGMSWITQGLLGALLLWMGINARDS	766
Qy	481	IAMTFLAVGGVLLFLSVNVHA	501
Db	767	IAMTFLAVGGVLLFLSVNVHA	787

RESULT 6

POLG WNV

ID POLG WNV STANDARD; PRT; 3430 AA.

AC P06935;

DT 01-JAN-1988, integrated into UniProtKB/Swiss-Prot.
 DT 24-OCT-2003, sequence version 2.
 DT 07-MAR-2006, entry version 64.
 DE Genome polyprotein [Contains: Capsid protein C (Core protein);
 DE Envelope protein M (Matrix protein); Major envelope protein E;
 DE Nonstructural protein 1 (NS1); Nonstructural protein 2A (NS2A);
 DE Flavivirin protease NS2B regulatory subunit; Flavivirin protease NS3
 DE catalytic subunit (EC 3.4.21.91); Nonstructural protein 4A (NS4A);
 DE Nonstructural protein 4B (NS4B); RNA-directed RNA polymerase
 DE (EC 2.7.7.48) (NS5)].
 OS West Nile virus (WN).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Flavivirus; Japanese encephalitis virus group.
 OX NCBI_TaxID=11082;
 RN [1]
 RP NUCLEOTIDE SEQUENCE [GENOMIC RNA].
 RX MEDLINE=86124703; PubMed=3753811;
 RA Castle E., Leidner U., Nowak T., Wengler G., Wengler G.;
 RT "Primary structure of the West Nile flavivirus genome region coding
 RT for all nonstructural proteins.";
 RL Virology 149:10-26(1986).
 RN [2]
 RP SEQUENCE REVISION TO 1908; 2018-2036; 2242 AND 2859-2860.
 RX MEDLINE=21176376; PubMed=11277701; DOI=10.1006/viro.2000.0795;
 RA Yamshchikov V.F., Wengler G., Perelygin A.A., Brinton M.A.,
 RA Compans R.W.;
 RT "An infectious clone of the West Nile flavivirus.";
 RL Virology 281:294-304(2001).
 RN [3]
 RP NUCLEOTIDE SEQUENCE [GENOMIC RNA] OF 1-291.
 RX MEDLINE=85274372; PubMed=2992152;
 RA Castle E., Nowak T., Leidner U., Wengler G., Wengler G.;
 RT "Sequence analysis of the viral core protein and the membrane-
 RT associated proteins V1 and NV2 of the flavivirus West Nile virus and
 RT of the genome sequence for these proteins.";
 RL Virology 145:227-236(1985).
 RN [4]
 RP NUCLEOTIDE SEQUENCE [GENOMIC RNA] OF 255-854.
 RX MEDLINE=86072082; PubMed=3855247;
 RA Wengler G., Castle E., Leidner U., Nowak T., Wengler G.;
 RT "Sequence analysis of the membrane protein V3 of the flavivirus West
 RT Nile virus and of its gene.";
 RL Virology 147:264-274(1985).
 RN [5]
 RP DISULFIDE BONDS IN E PROTEIN.
 RX MEDLINE=87122143; PubMed=3811228;
 RA Nowak T., Wengler G.;
 RT "Analysis of disulfides present in the membrane proteins of the West
 RT Nile flavivirus.";
 RL Virology 156:127-137(1987).
 CC -!- FUNCTION: The small proteins NS2A, NS4A and NS4B are hydrophobic,
 CC suggesting a possible membrane-related function. NS5 may play a
 CC role in the viral RNA replication. The NS2B/NS3 protease complex
 CC processes the viral polyprotein.
 CC -!- CATALYTIC ACTIVITY: Selective hydrolysis of -Xaa-Xaa-|-Yaa- bonds
 CC in which each of the Xaa can be either Arg or Lys and Yaa can be
 CC either Ser or Ala.

CC -!- CATALYTIC ACTIVITY: Nucleoside triphosphate + RNA(n) = diphosphate
 CC + RNA(n+1).
 CC -!- SUBUNIT: NS3 and NS2B form a heterodimer. NS3 is the catalytic
 CC subunit, whereas NS2B strongly stimulates the latter (By
 CC similarity).
 CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins
 CC (By similarity).
 CC -!- MISCELLANEOUS: The virion of this virus is a nucleocapsid covered
 CC by a lipoprotein envelope. The envelope contains two proteins: the
 CC protein M and glycoprotein E. The nucleocapsid is a complex of
 CC protein C and mRNA. In immature particles, there are 60
 CC icosaedrally organized trimeric spikes on the surface. Each spike
 CC consists of three heterodimers of envelope protein M precursor
 CC (prM) and envelope protein E (By similarity).
 CC -!- SIMILARITY: Contains 1 peptidase S7 domain.
 CC -!- SIMILARITY: Contains 1 RdRp catalytic domain.

CC -----
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License
 CC -----

DR EMBL; M12294; AAA48498.2; -; Genomic_RNA.
 DR PIR; A25256; GNWVWV.
 DR HSSP; Q88653; 1L9K.
 DR SMR; P06935; 25-97.
 DR MEROPS; S07.001; -.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR011545; DEAD/DEAH_N.
 DR InterPro; IPR002464; DEAH_box.
 DR InterPro; IPR011999; Flav_glyE_cen_dm.
 DR InterPro; IPR001122; Flavi_capsidC.
 DR InterPro; IPR011492; Flavi_DEAD.
 DR InterPro; IPR000069; Flavi_M.
 DR InterPro; IPR001157; Flavi_NS1.
 DR InterPro; IPR000752; Flavi_NS2A.
 DR InterPro; IPR000487; Flavi_NS2B.
 DR InterPro; IPR000404; Flavi_NS4A.
 DR InterPro; IPR001528; Flavi_NS4B.
 DR InterPro; IPR000208; Flavi_NS5.
 DR InterPro; IPR002535; Flavi_propep.
 DR InterPro; IPR000336; Flv_glyE_Ig-like.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR001850; Peptidase_S7.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR InterPro; IPR002877; RrmJFtsJ_mtfrase.
 DR InterPro; IPR011998; Vrl_glyE_cen_dim.
 DR Pfam; PF01003; Flavi_capsid; 1.
 DR Pfam; PF07652; Flavi_DEAD; 1.
 DR Pfam; PF02832; Flavi_glycop_C; 1.
 DR Pfam; PF00869; Flavi_glycoprot; 1.
 DR Pfam; PF01004; Flavi_M; 1.
 DR Pfam; PF00948; Flavi_NS1; 1.
 DR Pfam; PF01005; Flavi_NS2A; 1.
 DR Pfam; PF01002; Flavi_NS2B; 1.
 DR Pfam; PF01350; Flavi_NS4A; 1.
 DR Pfam; PF01349; Flavi_NS4B; 1.
 DR Pfam; PF00972; Flavi_NS5; 1.

DR Pfam; PF01570; Flavi_propep; 1.
 DR Pfam; PF01728; FtsJ; 1.
 DR Pfam; PF00271; Helicase_C; 1.
 DR Pfam; PF00949; Peptidase_S7; 1.
 DR ProDom; PD001496; Flavi_NS1; 1.
 DR SMART; SM00487; DEXDc; 1.
 DR SMART; SM00490; HELICc; 1.
 DR PROSITE; PS00690; DEAH_ATP_HELICASE; FALSE_NEG.
 DR PROSITE; PS50507; RDRP_SSRNA_POS; 1.
 KW ATP-binding; Capsid protein; Core protein; Envelope protein;
 KW Glycoprotein; Helicase; Hydrolase; Membrane; Nucleotide-binding;
 KW Nucleotidyltransferase; Polyprotein; RNA-directed RNA polymerase;
 KW Structural protein; Transferase; Transmembrane.
 FT CHAIN 1 123 Capsid protein C.
 FT /FTId=PRO_0000037743.
 FT INIT_MET 1 1 Removed from capsid protein C by the
 FT cellular aminopeptidase.
 FT PROPEP 124 215
 FT /FTId=PRO_0000037744.
 FT CHAIN 216 290 Envelope protein M.
 FT /FTId=PRO_0000037745.
 FT CHAIN 291 787 Major envelope protein E.
 FT /FTId=PRO_0000037746.
 FT CHAIN 788 1139 Nonstructural protein 1.
 FT /FTId=PRO_0000037747.
 FT CHAIN 1140 1370 Nonstructural protein 2A.
 FT /FTId=PRO_0000037748.
 FT CHAIN 1371 1501 Flavivirin protease NS2B regulatory
 FT subunit.
 FT /FTId=PRO_0000037749.
 FT CHAIN 1502 2120 Flavivirin protease NS3 catalytic
 FT subunit.
 FT /FTId=PRO_0000037750.
 FT CHAIN 2121 2269 Nonstructural protein 4A.
 FT /FTId=PRO_0000037751.
 FT CHAIN 2270 2525 Nonstructural protein 4B.
 FT /FTId=PRO_0000037752.
 FT CHAIN 2526 3430 RNA-directed RNA polymerase.
 FT /FTId=PRO_0000037753.
 FT DOMAIN 1508 1679 Peptidase S7.
 FT DOMAIN 3055 3207 RdRp catalytic.
 FT NP_BIND 1695 1702 ATP (Potential).
 FT REGION 388 401 Involved in fusion.
 FT MOTIF 1786 1789 DEAH box.
 FT ACT_SITE 1552 1552 Charge relay system (By similarity).
 FT ACT_SITE 1576 1576 Charge relay system (By similarity).
 FT ACT_SITE 1636 1636 Charge relay system (By similarity).
 FT CARBOHYD 138 138 N-linked (GlcNAc . . .) (Potential).
 FT CARBOHYD 917 917 N-linked (GlcNAc . . .) (Potential).
 FT CARBOHYD 962 962 N-linked (GlcNAc . . .) (Potential).
 FT CARBOHYD 994 994 N-linked (GlcNAc . . .) (Potential).
 FT CARBOHYD 1289 1289 N-linked (GlcNAc . . .) (Potential).
 FT CARBOHYD 2336 2336 N-linked (GlcNAc . . .) (Potential).
 FT CARBOHYD 2489 2489 N-linked (GlcNAc . . .) (Potential).
 FT DISULFID 293 320
 FT DISULFID 350 406
 FT DISULFID 364 395

FT DISULFID 382 411
 FT DISULFID 476 574
 FT DISULFID 591 622
 SQ SEQUENCE 3430 AA; 380110 MW; 42D71B7CB12DC45B CRC64;

Query Match 97.5%; Score 2572; DB 1; Length 3430;
 Best Local Similarity 98.2%; Pred. No. 2.9e-186;
 Matches 492; Conservative 2; Mismatches 3; Indels 4; Gaps 1;

```

Qy      1 FNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVKMMNMEAANLADVRSYC 60
      |||
Db     291 FNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVKMMNMEAANLADVRSYC 350

Qy      61 YLASVSDLSTKAACPTMGEAHNEKRA DPAFVCKQGVVDRGWNGCGLFGKGSIDTCAKFA 120
      |||:|||||
Db     351 YLASVSDLSTRAACPTMGEAHNEKRA DPAFVCKQGVVDRGWNGCGLFGKGSIDTCAKFA 410

Qy     121 CTTKATGWIIQKENIKYEVAIFVHGPTTVESHGNYSTQIGATQAGRFSITPSAPSYTLKL 180
      |||:|||||
Db     411 CTTKATGWIIQKENIKYEVAIFVHGPTTVESHG----KIGATQAGRFSITPSAPSYTLKL 466

Qy     181 GEYGEVTV DCEPRSGIDTSAYYVMSVGAKSFLVHREWFMDLNL PWSSAGSTTWRNRETLM 240
      |||
Db     467 GEYGEVTV DCEPRSGIDTSAYYVMSVG EKSFLVHREWFMDLNL PWSSAGSTTWRNRETLM 526

Qy     241 EFEEPHATKQSVVALGSQEGALHQALAGAI PVEFSSNTVKLTSGHLKCRVKMEKLQLKGT 300
      |||
Db     527 EFEEPHATKQSVVALGSQEGALHQALAGAI PVEFSSNTVKLTSGHLKCRVKMEKLQLKGT 586

Qy     301 TYGVCSKAFKFARTPADTGHGTVVLELQYTGKDGPCKVPISSVASLNDLTPVGR LVTVNP 360
      |||
Db     587 TYGVCSKAFKFARTPADTGHGTVVLELQYTG DGPCKVPISSVASLNDLTPVGR LVTVNP 646

Qy     361 FVSVATANSKVLIELEPPFSDSYIVVGRGEQQINHHWHKSGSSIGKAFTTTLRGAQRLAA 420
      |||
Db     647 FVSVATANSKVLIELEPPFGDSYIVVGRGE QQINHHWHKSGSSIGKAFTTTLRGAQRLAA 706

Qy     421 LGDTAWDFGSGGVFTSVGKAIHQVFGGAFRSLF GGM SWITQGLLGALLLW MGINARDRS 480
      |||
Db     707 LGDTAWDFGSGGVFTSVGKAIHQVFGGAFR SLFGGM SWITQGLLGALLLW MGINARDRS 766

Qy     481 IAMTFLAVGGVLLFLSVNVHA 501
      |||
Db     767 IAMTFLAVGGVLLFLSVNVHA 787

```

RESULT 7

Q5EVN3_WNV

ID Q5EVN3_WNV PRELIMINARY; PRT; 3433 AA.

AC Q5EVN3;

DT 15-MAR-2005, integrated into UniProtKB/TrEMBL.

DT 15-MAR-2005, sequence version 1.

DT 07-FEB-2006, entry version 5.

DE Polyprotein.

OS West Nile virus (WN).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Flavivirus; Japanese encephalitis virus group.

OX NCBI_TaxID=11082;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=96-111;
 RX PubMed=15752452;
 RA Schuffenecker I., Peyrefitte C.N., el Harrak M., Murri S., Leblond A.,
 RA Zeller H.G.;
 RT "West Nile Virus in Morocco, 2003.";
 RL Emerg. Infect. Dis. 11:306-309(2005).
 CC -----
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License
 CC -----
 DR EMBL; AY701412; AAT92098.1; -; Genomic_RNA.
 DR SMR; Q5EVN3; 25-97.
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0019031; C:viral envelope; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
 DR GO; GO:0003725; F:double-stranded RNA binding; IEA.
 DR GO; GO:0003724; F:RNA helicase activity; IEA.
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
 DR GO; GO:0004252; F:serine-type endopeptidase activity; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0019079; P:viral genome replication; IEA.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR011545; DEAD/DEAH_N.
 DR InterPro; IPR011999; Flav_glyE_cen_dm.
 DR InterPro; IPR001122; Flavi_capsidC.
 DR InterPro; IPR011492; Flavi_DEAD.
 DR InterPro; IPR000069; Flavi_M.
 DR InterPro; IPR001157; Flavi_NS1.
 DR InterPro; IPR000752; Flavi_NS2A.
 DR InterPro; IPR000487; Flavi_NS2B.
 DR InterPro; IPR000404; Flavi_NS4A.
 DR InterPro; IPR001528; Flavi_NS4B.
 DR InterPro; IPR000208; Flavi_NS5.
 DR InterPro; IPR002535; Flavi_propep.
 DR InterPro; IPR000336; Flv_glyE_Ig-like.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR001850; Peptidase_S7.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSvir.
 DR InterPro; IPR002877; RrmJFtsJ_mtfase.
 DR InterPro; IPR011998; Vrl_glyE_cen_dim.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF01003; Flavi_capsid; 1.
 DR Pfam; PF07652; Flavi_DEAD; 1.
 DR Pfam; PF02832; Flavi_glycop_C; 1.
 DR Pfam; PF00869; Flavi_glycoprot; 1.
 DR Pfam; PF01004; Flavi_M; 1.
 DR Pfam; PF00948; Flavi_NS1; 1.
 DR Pfam; PF01005; Flavi_NS2A; 1.
 DR Pfam; PF01002; Flavi_NS2B; 1.
 DR Pfam; PF01350; Flavi_NS4A; 1.
 DR Pfam; PF01349; Flavi_NS4B; 1.

DR Pfam; PF00972; Flavi_NS5; 1.
 DR Pfam; PF01570; Flavi_propep; 1.
 DR Pfam; PF01728; FtsJ; 1.
 DR Pfam; PF00271; Helicase_C; 1.
 DR Pfam; PF00949; Peptidase_S7; 1.
 DR ProDom; PD001496; Flavi_NS1; 1.
 DR SMART; SM00487; DEXDc; 1.
 DR SMART; SM00490; HELICc; 1.
 DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1..
 KW Polyprotein.
 SQ SEQUENCE 3433 AA; 381249 MW; 7ECC96DBFD9D53DA CRC64;

Query Match 96.0%; Score 2532; DB 2; Length 3433;
 Best Local Similarity 95.6%; Pred. No. 3.2e-183;
 Matches 479; Conservative 13; Mismatches 9; Indels 0; Gaps 0;

```

Qy      1  FNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVKMMNMEAANLADVRSYC 60
          |||:|||||
Db      291 FNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVKMMNMEAANLAEVRSYC 350

Qy      61  YLASVSDLSTKAACPTMGEAHNEKRADPAFVCKQGVVDRGWNGCGLFGKGSIDTCAKFA 120
          |||:|||||
Db      351 YLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWNGCGLFGKGSIDTCAKFA 410

Qy      121 CTTKATGWIIQKENIKYEVAIFVHGPTTVESHGNYSTQIGATQAGRFSITPSAPSYTLKL 180
          |:|||||
Db      411 CSTKATGRTILKENIKYEVAIFVHGPTTVESHGNYSTQIGATQAGRFSITPAAPSYTLKL 470

Qy      181 GEYGEVTVDCEPRSGIDTSAYYVMSVGAKSFLVHREWFMDLNLPWSSAGSTTWRNRETLM 240
          |||:|||||
Db      471 GEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWFMDLNLPWSSAGSTVWRNRETLM 530

Qy      241 EFEEPHATKQSVVALGSQEGALHQALAGAIPEFSSNTVKLTSGHLKCRVKMEKLQLKGT 300
          |||:|||||
Db      531 EFEEPHATKQSVIALGSQEGALHQALAGAIPEFSSNTVKLTSGHLKCRVKMEKLQLKGT 590

Qy      301 TYGVCSKAFKFARTPADTGHGTVVLELQYTGKDGPKVPISSVASLNDLTPVGRLVTVNP 360
          |||:|||||
Db      591 TYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDGPKVPISSVASLNDLTPVGRLVTVNP 650

Qy      361 FVSVATANSKVLIELEPPFSDSYIVVGRGEQQINHHWHKSGSSIGKAFTTTLRGAQRLAA 420
          |||:|||||
Db      651 FVSVATANAKVLIELEPPFGDSYIVVGRGEQQINHHWHKSGSSIGKAFTTTLKGAQRLAA 710

Qy      421 LGDTAWDFGSVGGVFTSVGKAIHQVFGAFRSLFGGMSWITQGLLGALLLWMGINARDS 480
          |||:|||||
Db      711 LGDTAWDFGSVGGVFTSVGKAIHQVFGAFRSLFGGMSWITQGLLGALLLWMGINARDS 770

Qy      481 IAMTFLAVGGVLLFLSVNVHA 501
          ||:|||||
Db      771 IALTFLAVGGVLLFLSVNVHA 791

```

RESULT 8

Q9WI84_WNV

ID Q9WI84_WNV PRELIMINARY; PRT; 501 AA.

AC Q9WI84;

DR GO; GO:0019031; C:viral envelope; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0019058; P:viral infectious cycle; IEA.
 DR InterPro; IPR011999; Flav_glyE_cen_dm.
 DR InterPro; IPR001122; Flavi_capsidC.
 DR InterPro; IPR000069; Flavi_M.
 DR InterPro; IPR002535; Flavi_propep.
 DR InterPro; IPR000336; Flv_glyE_Ig-like.
 DR InterPro; IPR011998; Vrl_glyE_cen_dim.
 DR Pfam; PF01003; Flavi_capsid; 1.
 DR Pfam; PF02832; Flavi_glycop_C; 1.
 DR Pfam; PF00869; Flavi_glycoprot; 1.
 DR Pfam; PF01004; Flavi_M; 1.
 DR Pfam; PF01570; Flavi_propep; 1.
 KW Polyprotein.
 FT CHAIN <1 88 capsid protein.
 FT CHAIN 89 265 pre-membrane/membrane protein.
 FT CHAIN 266 766 envelope glycoprotein.
 FT NON_TER 1 1
 FT NON_TER 773 773
 SQ SEQUENCE 773 AA; 83364 MW; 2C33EA27EC676EE7 CRC64;

Query Match 95.9%; Score 2531; DB 2; Length 773;
 Best Local Similarity 95.4%; Pred. No. 5.1e-184;
 Matches 478; Conservative 14; Mismatches 9; Indels 0; Gaps 0;

Qy 1 FNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVKMMNMEAANLADVRSYC 60
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||
 Db 266 FNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVKMMNMEAANLAEVRSYC 325
 Qy 61 YLASVSDLSTKAACPTMGEAHNEKRAPAFVCKQGVVDRGWNGCGLFGKGSIDTCAKFA 120
 |||:|||||||||||||||||:|||||||:|||||||||||||||||
 Db 326 YLATVSDLSTKAACPTMGEAHNDRAPAFVCRQGVVDRGWNGCGLFGKGSIDTCAKFA 385
 Qy 121 CTTKATGWIIQKENIKYEVAIFVHGPTTVESHGNYSTQIGATQAGRFSITPSAPSYTLKL 180
 |:|||| | |||||||||||||||||||||||||||||||||:|||||||
 Db 386 CSTKATGRTILKENIKYEVAIFVHGPTTVESHGNYSTQIGATQAGRFSITPAAPSYTLKL 445
 Qy 181 GEYGEVTVDCPRSGIDTSAYYVMSVGAKSFLVHREWFMDLNLPWSSAGSTTWRNRETLM 240
 |||||||||||||||:|||||:| |:||||||||| |||||||
 Db 446 GEYGEVTVDCPRSGIDTNAYYVMTVGTKTFLVHREWFMDLNLPWSSAGSTVWRNRETLM 505
 Qy 241 EFEEPHATKQSVVALGSQEGALHQALAGAIPEFSSNTVKLTSGHLKCRVKMEKLQLKGT 300
 |||||||||||:|||||||||||||||||||||||||||||||
 Db 506 EFEEPHATKQSVIALGSQEGALHQALAGAIPEFSSNTVKLTSGHLKCRVKMEKLQLKGT 565
 Qy 301 TYGVCSKAFKFARTPADTGHGTVVLELQYTGKDGPKVPISSVASLNDLTPVGRLVTNP 360
 ||||||||| ||||||||||||||| |||||||||||||
 Db 566 TYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDGPKVPISSVASLNDLTPVGRLVTNP 625
 Qy 361 FVSVATANSKVLIELEPPFSDSYIVVGRGEQQINHHWHKSGSSIGKAFTTTLRGAQRLAA 420
 |||||||:||||||| ||||||||||||||| |||||||||:|||||||
 Db 626 FVSVATANAKVLIELEPPFGDSYIVVGRGEQQINHHWHKSGSSIGKAFTTTLKGAQRLAA 685
 Qy 421 LGDTAWDFGSGGVFTSVGKAIHQVFGGAFRSLFGGMSWITQGLLGALLLWMGINARDS 480
 |||||||||||||||:||||||| |||||||||||||
 Db 686 LGDTAWDFGSGGVFTSVGKAVHQVFGGAFRSLFGGMSWITQGLLGALLLWMGINARDS 745

Qy 481 IAMTFLAVGGVLLFLSVNVHA 501
 ||:|||||
 Db 746 IALTFLAVGGVLLFLSVNVHA 766

RESULT 10

Q5EVN2_WNV

ID Q5EVN2_WNV PRELIMINARY; PRT; 3433 AA.
 AC Q5EVN2;
 DT 15-MAR-2005, integrated into UniProtKB/TrEMBL.
 DT 15-MAR-2005, sequence version 1.
 DT 07-FEB-2006, entry version 5.
 DE Polyprotein.
 OS West Nile virus (WN).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Flavivirus; Japanese encephalitis virus group.
 OX NCBI_TaxID=11082;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=04.05;
 RX PubMed=15752452;
 RA Schuffenecker I., Peyrefitte C.N., el Harrak M., Murri S., Leblond A.,
 RA Zeller H.G.;
 RT "West Nile Virus in Morocco, 2003.";
 RL Emerg. Infect. Dis. 11:306-309(2005).
 CC -----
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License
 CC -----
 DR EMBL; AY701413; AAT92099.1; -; Genomic_RNA.
 DR SMR; Q5EVN2; 25-97.
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0019031; C:viral envelope; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
 DR GO; GO:0003725; F:double-stranded RNA binding; IEA.
 DR GO; GO:0003724; F:RNA helicase activity; IEA.
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
 DR GO; GO:0004252; F:serine-type endopeptidase activity; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0019079; P:viral genome replication; IEA.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR011545; DEAD/DEAH_N.
 DR InterPro; IPR011999; Flav_glyE_cen_dm.
 DR InterPro; IPR001122; Flavi_capsidC.
 DR InterPro; IPR011492; Flavi_DEAD.
 DR InterPro; IPR000069; Flavi_M.
 DR InterPro; IPR001157; Flavi_NS1.
 DR InterPro; IPR000752; Flavi_NS2A.
 DR InterPro; IPR000487; Flavi_NS2B.
 DR InterPro; IPR000404; Flavi_NS4A.
 DR InterPro; IPR001528; Flavi_NS4B.
 DR InterPro; IPR000208; Flavi_NS5.
 DR InterPro; IPR002535; Flavi_propep.
 DR InterPro; IPR000336; Flv_glyE_Ig-like.

Qy 361 FVSVATANSKVLIELEPPFSDSYIVVGRGEQQINHHWHKSGSSIGKAFTTTTLRGAQRLAA 420
 |||||:||||| |||||:|||||
 Db 651 FVSVATANAKVLIELEPPFGDSYIVVGRGEQQINHHWHKSGSSIGKAFTTTTLKGAQRLAA 710
 Qy 421 LGDTAWDFGSVGGVFTSVGKAIHQVFGGAFRSLFGGMSWITQGLLGALLLWMGINARDS 480
 |||||:||||| |||||:|||||
 Db 711 LGDTAWDFGSVGGVFTSVGKAVHQVFGGAFRSLFGGMSWITQGLLGALLLWMGINARDS 770
 Qy 481 IAMTFLAVGGVLLFLSVNVHA 501
 ||:|||||
 Db 771 IALTFLAVGGVLLFLSVNVHA 791

RESULT 11

Q6WV07_WNV

ID Q6WV07_WNV PRELIMINARY; PRT; 3433 AA.
 AC Q6WV07;
 DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.
 DT 05-JUL-2004, sequence version 1.
 DT 07-FEB-2006, entry version 8.
 DE Polyprotein.
 GN Name=pol;
 OS West Nile virus (WN).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Flavivirus; Japanese encephalitis virus group.
 OX NCBI_TaxID=11082;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=PaAn001;
 RX MEDLINE=22949215; PubMed=14585341; DOI=10.1016/S0042-6822(03)00536-1;
 RA Charrel R.N., Brault A.C., Gallian P., Lemasson J.-J., Murgue B.,
 RA Murri S., Pastorino B., Zeller H., de chesse R., de Micco P.,
 RA de Lamballerie X.;
 RT "Evolutionary relationship between Old World West Nile virus strains.
 RT Evidence for viral gene flow between Africa, the Middle East, and
 RT Europe."
 RL Virology 315:381-388(2003).
 CC -----
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License
 CC -----
 DR EMBL; AY268132; AAQ00998.1; -; Genomic_RNA.
 DR HSSP; Q9Q4T1; 1BEF.
 DR SMR; Q6WV07; 25-97.
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0019031; C:viral envelope; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
 DR GO; GO:0003725; F:double-stranded RNA binding; IEA.
 DR GO; GO:0003724; F:RNA helicase activity; IEA.
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
 DR GO; GO:0004252; F:serine-type endopeptidase activity; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0019079; P:viral genome replication; IEA.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR011545; DEAD/DEAH_N.

Qy	181	GEYGEVTVDCEPRSGIDTSAYYVMSVGA	KSFLVHREWFMDLNL	PWSSAGSTTWRNRETLM	240
		: :	:		
Db	471	GEYGEVTVDCEPRSGIDTNAYYVMTVGT	KTFLVHREWFMDLNL	PWSSAGSTVWRNRETLM	530
Qy	241	EFEEPHATKQSVVALG	SQEGALHQALAGAI	PVEFSSNTVKLTSGHLKCRVKMEKLQ	300
		:			
Db	531	EFEEPHATKQSVIALG	SQEGALHQALAGAI	PVEFSSNTVKLTSGHLKCRVKMEKLQ	590
Qy	301	TYGVCSKAFKFARTPADTGHGTVVLELQ	YTGKDGPKVPISSVASLNDLTPVGR	LVTVNP	360
Db	591	TYGVCSKAFKFLGTPADTGHGTVVLELQ	YTGTDGPKVPISSVASLNDLTPVGR	LVTVNP	650
Qy	361	FVSVATANSKVLIELEPPFSDSYIVVGR	GEQQINHHWHKSGSSIGKAFTTT	LGAQRLAA	420
		:		:	
Db	651	FVSVATANAKVLIELEPPFGDSYIVVGR	GEQQINHHWHKSGSSIGKAFTTT	LKAQRLAA	710
Qy	421	LGDTAWDFGSVGGVFTSVGKAIHQVFGG	AFRSLFGGMSWITQGLLGALLLW	MGINARDS	480
		:			
Db	711	LGDTAWDFGSVGGVFTSVGKAHVHQVFGG	AFRSLFGGMSWITQGLLGALLLW	MGINARDS	770
Qy	481	IAMTFLAVGGVLLFLSVNVHA	501		
		:			
Db	771	IALTFLAVGGVLLFLSVNVHA	791		